3-14-2014

# Impact of the Shodan Computer Search Engine on Internet-facing Industrial Control System Devices

Roland C. Bodenheim


# IMPACT OF THE SHODAN COMPUTER SEARCH ENGINE ON INTERNET-FACING INDUSTRIAL CONTROL SYSTEM DEVICES

THESIS

Roland C. Bodenheim, Captain, USAF

AFIT-ENG-14-M-14

**DEPARTMENT OF THE AIR FORCE**
**AIR UNIVERSITY**

***AIR FORCE INSTITUTE OF TECHNOLOGY***

**Wright-Patterson Air Force Base, Ohio**

AFIT-ENG-14-M-14

# IMPACT OF THE SHODAN COMPUTER SEARCH ENGINE ON INTERNET-FACING INDUSTRIAL CONTROL SYSTEM DEVICES

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cyber Operations

Roland C. Bodenheim, B.S.C.S.

Captain, USAF

March 2014

AFIT-ENG-14-M-14

# IMPACT OF THE SHODAN COMPUTER SEARCH ENGINE ON INTERNET-FACING INDUSTRIAL CONTROL SYSTEM DEVICES

Roland C. Bodenheim, B.S.C.S.
Captain, USAF

Approved:

| | |
|---|---|
| //signed// | 14 March 2014 |
| Maj Jonathan Butts, PhD (Chairman) | Date |
| //signed// | 14 March 2014 |
| Barry Mullins, PhD (Member) | Date |
| //signed// | 14 March 2014 |
| Stephen Dunlap, MS (Member) | Date |

AFIT-ENG-14-M-14


## Abstract

The Shodan computer search engine crawls the Internet attempting to identify any connected device. Using Shodan, researchers identified thousands of Internet-facing devices associated with industrial controls systems (ICS). This research examines the impact of Shodan on ICS security, evaluating Shodans ability to identify Internet-connected ICS devices and assess if targeted attacks occur as a result of Shodan identification. In addition, this research evaluates the ability to limit device exposure to Shodan through service banner manipulation. Shodans impact was evaluated by deploying four high-interaction, unsolicited honeypots over a 55 day period, each configured to represent Allen-Bradley programmable logic controllers (PLC). All four honeypots were successfully indexed and identifiable via the Shodan web interface in less than 19 days. Despite being indexed, there was no increased network activity or targeted ICS attacks. Although results indicate Shodan is an effective reconnaissance tool, results contrast claims of its use to broadly identify and target Internet-facing ICS devices. Additionally, the service banner for two PLCs were modified to evaluate the impact on Shodan indexing capabilities. Findings demonstrated service banner manipulation successfully limited device exposure from Shodan queries.

## Table of Contents

Page

Page

## List of Figures

## List of Tables

Table Page

## List of Acronyms

| Acronym | Definition |
|---|---|
| API | application programming interface |
| CCTV | closed circuit television |
| CIA | Central Intelligence Agency |
| CIAG | Critical Infrastructure Assurance Group |
| CIP | common industrial protocol |
| CPU | central processing unit |
| CTO | Chief Technology Officer |
| DCS | distributed control system |
| DHS | Department of Homeland Security |
| DMZ | demilitarized zones |
| DNP3 | distributed network protocol |
| FTP | file transfer protocol |
| GAO | Government Accountability Office |
| HMI | human machine interface |
| HTTP | hypertext transfer protocol |
| ICS | industrial control system |
| ICS-CERT | Industrial Control System Cyber Emergency Response Team |
| INL | Idaho National Laboratory |
| IDS | intrusion detection system |
| IEEE | Institute of Electrical and Electronics Engineers |
| EtherNet/IP | Ethernet industrial protocol |
| IED | intelligent electronic device |
| IP | Internet protocol |

| Acronym | Definition |
|---|---|
| IPv4 | Internet protocol version 4 |
| IT | information technology |
| MITM | man-in-the-middle |
| MTU | master terminal unit |
| NIST | National Institute of Standards and Technology |
| NMAP | network mapper |
| NRC | National Research Council |
| OS | operating system |
| ODVA | Open DeviceNet Vendors Association |
| PLC | programmable logic controller |
| RTU | remote terminal unit |
| SCADA | supervisory control and data acquisition |
| SNMP | simple network management protocol |
| UDP | user datagram protocol |
| US-CERT | United States Computer Emergency Response Team |
| TCP | transmission control protocol |
| TCP/IP | transmission control protocol/Internet protocol |
| VM | virtual machine |

IMPACT OF THE SHODAN COMPUTER SEARCH ENGINE ON
INTERNET-FACING INDUSTRIAL CONTROL SYSTEM DEVICES

# I. Introduction

## 1.1 Background

Industrial control systems (ICS) are integral to United States critical infrastructure, allowing real-time remote management of large-scale industrial processing supporting oil and gas pipelines, water distribution systems, electrical power grids, nuclear plants, and other manufacturing operations. In 2005, the SANS Institute estimated over 3 million active ICSs, with an expected 8.9% annual growth culminating in approximately 6 million ICSs by 2013, with nearly all critical infrastructure sectors moving to advanced control systems [23]. This growth, combined with market demand, has lead to a shift towards ICS network connectivity to lower operational costs and increase efficiency. In some cases ICSs are connected to the corporate networks, while in other more pernicious circumstances ICSs are directly accessible via the Internet.

In fiscal 2012, Industrial Control System Cyber Emergency Response Team (ICS-CERT) responded to 198 cyber incidents involving critical infrastructure systems, a 65% increase over the 120 attacks reported in 2011 [26]. In addition, recent research identified thousands of ICS associated devices readily accessible via the Internet [6, 30]. The steady rise in cyber incidents combined with exponential growth and increased connectivity presents a monumental security risk to United States national security.

## 1.2 Motivation

In 2009, John Matherly launched Shodan, a computer search engine designed to identify and index Internet-facing devices [43]. Four years later, CNN referred to Shodan as "The scariest search engine on the Internet," reporting that Shodan collects information on more than 500 million devices and services a month [18].

Shodan is a search engine that scans the Internet for any Internet-facing device. The Shodan database contains web and security cameras, home automation, traffic lights, car washes, and even an entire hockey rink [18]. One of the most discerning aspects is the vast number of industrial control devices identifiable via Shodan. These devices control critical infrastructure to include oil and gas pipelines, water, power grids, and nuclear plants. Department of Homeland Security (DHS) stated Shodan allows malicious and skilled adversaries ready access to admittedly fragile systems, some of which support United States critical infrastructure [39]. Since Shodan's launch in 2009, direct Internet ICS connectivity has continued to grow despite insistent urgings from public and private security experts. Research has demonstrated that Shodan is a capable reconnaissance tool [6, 30] and has shown Internet-facing ICS devices are being attacked [52]; however, there lacks empirical evidence to support the claim Shodan is actively being used to target Internet-facing ICS devices and if it is being used, what the impact is on ICS device security.

## 1.3 Problem Statement

This research evaluates Shodan's impact on Internet-facing ICS device security. The primary goals of this research are to evaluate Shodan indexing functionality, contrast network activity levels as a result of Shodan indexing and identification, and enumerate ICS specific targeting of Internet-facing ICS devices. The secondary goal is to assess the ability to limit Shodan device exposure via service banner manipulation. It is hypothesized that as a result of Shodan identification, Internet-facing ICS devices will see

increased network activity, to include specific attacks targeting ICS protocols. It is further hypothesized that a device presenting a more enticing service banner will see increased network activity post Shodan identification as compared to both the standard and obfuscated devices.

## 1.4 Approach

This research presents an evaluation of the Shodan computer search engine's impact on Internet-facing ICS device security by deploying unsolicited, Internet-facing ICS honeypots. The honeypots are designed and configured to be representative of ICS devices currently identifiable via Shodan. The primary goals are to evaluate Shodan indexing functionality, contrast network activity levels as a result of Shodan identification, and enumerate specific ICS targeting or attacks. Shodan indexing functionality is evaluated by determining Shodan's scanning routine, scanning frequency, and web database identification timeliness. Network activity is analyzed by measuring transmission control protocol (TCP) connections, total TCP packet count, and the number of unique Internet protocol (IP) addresses interacting with each honeypot. Shodan identification is defined as the date a device service banner is successfully indexed and the device is identifiable via the Shodan web interface. Specific ICS targeting and attacks are evaluated by visual packet inspection and analysis using the Snort intrusion detection system (IDS) with known ICS signatures.

The cornerstone of Shodan is a database containing Internet-facing device service banners. A service banner refers to information provided by a system in response to a connection request. Using the information revealed in a service banner, users are able to craft custom search queries in Shodan capable of specifically identifying ICS devices. To assess the ability to limit Shodan device exposure via service banner manipulation, two honeypots are deployed with altered service banners. One honeypot presents an enticing

service banner directly identifying the device make and model, while the second honeypot replaces the service banner with random data in attempts to obfuscate the device.

The results of this research provide insight into Shodan's impact on Internet-facing ICS devices. Specifically, results indicate if Shodan is actively being used as a passive reconnaissance tool to target Internet-facing ICS devices. The results of the banner mangled honeypot analysis reveal potential defensive measures to limit a device's exposure to Shodan query identification.

## 1.5 Assumptions and Limitations

This research provides an indication of Shodan's impact on Internet-facing ICS device security. Previous research efforts provide evidence ICS devices are deployed Internet-facing [6], Shodan is capable of identifying these ICS devices [30], and Internet-facing ICS devices are being directly targeted [52]. No research, however, currently exists assessing the correlation between Shodan device identification and specific device targeting. This section presents the assumptions and limitations of the research. Future research can build on this work and address these limitations.

### *1.5.1 Scope.*

The scope of this research extends to Internet-facing ICS programmable logic controllers (PLCs), specifically Allen-Bradley ControlLogix PLCs. A characterization of Allen-Bradley ControlLogix PLCs currently identifiable via Shodan is used to develop the design for each honeypot. Future research should extend ICS honeypot deployment to measure the impact of Shodan on additional ICS field devices and device manufacturers.

### *1.5.2 Time.*

The primary goal of this research is to evaluate the impact of Shodan on Internet-facing ICS security. As such, the assessment evaluated a 55 day deployment period. The deployment period was selected based on previous ICS honeypot research and an approximation of the time required to scan all public Internet protocol version 4 (IPv4)

addresses. It is assumed the findings are indicative of Shodan characteristics. Future research could consider longer deployment to obtain a larger dataset for analysis.

#### *1.5.3 Programmable Logic Controller.*

This research uses the Allen-Bradley ControlLogix 1756-L61 central processing unit (CPU) module and eWeb Ethernet module. Honeypot design and configuration is based on a random sampling of 10% of Allen-Bradley ControlLogix PLCs currently identifiable via Shodan. Specifically, nine characteristics: CPU type, CPU firmware version, Ethernet module type, Ethernet module firmware version, naming conventions, number of modules, type of modules, chassis size, and available services. It is assumed the characterization of Allen-Bradley ControlLogix PLCs identifiable via Shodan allows the results to be representative of the larger population of Internet-facing Allen-Bradley ControlLogix PLCs. Future research may extend to additional PLC manufacturers and models.

#### *1.5.4 Deployment Location.*

The ability to obtain Internet-facing IP space co-located with an ICS entity limited the honeynet design and the size of the honeynet. Available resources allowed for the deployment of four ICS honeypots and dictated the honeypots be deployed with sequential static IP addresses in the same subnet. Although there was no evidence this impacted the results of this research, future research could seek a broader deployment to include multiple venues and additional honeypots.

### 1.6 Thesis Organization

Chapter 2 presents a detailed background and overview of related research. Chapter 3 provides the research methodology to include honeypot design, deployment, and data evaluation. Chapter 4 provides implementation details as well as the results of this research. Chapter 5 discusses the research conclusions, future work, and concluding remarks.

# II. Background

## 2.1 Industrial Control Systems

ICSs date back to 1959 with the deployment of a Thompson Ramo Wooldridge RW-300 direct digital control process computer installed at the Texaco refinery in Port Arthur, Texas [24]. ICSs encompass several different control systems utilized throughout the industrial processing and critical infrastructure communities. By design, these systems allow real-time remote management of large-scale industrial processing [3]. They monitor and control critical infrastructure supporting oil and gas pipelines, water distribution systems, electrical power grids, nuclear plants, and other manufacturing operations. Two types of ICSs are supervisory control and data acquisition (SCADA) systems and distributed control systems (DCSs).

SCADA systems are highly distributed systems, providing automated control and remote human monitoring of real-world processes [3]. The supervisory control aspect of SCADA relates to the operator's ability to control remote processes via field devices such as remote terminal units (RTUs) and PLCs. Data acquisition references the transfer of data from RTUs and PLCs to a centralized control center where it is displayed to the operator via a human machine interface (HMI). Figure 2.1 depicts a typical SCADA system architecture. SCADA system architecture is comprised of four layers. Layer one, consists of the physical ICS assets (e.g., mechanical valve or digital temperature gauge). Layer two is comprised of field devices (e.g., RTUs and PLCs). Layer three is made up of the control network housing the actual ICS. The uppermost layer, layer four, is home of the traditional information technology (IT) network, hosting the corporate network and controlling site manufacturing operations [8].

DCSs are computerized control systems wherein the controlling elements are not centralized [3]. Rather, control is distributed throughout the system, with each component

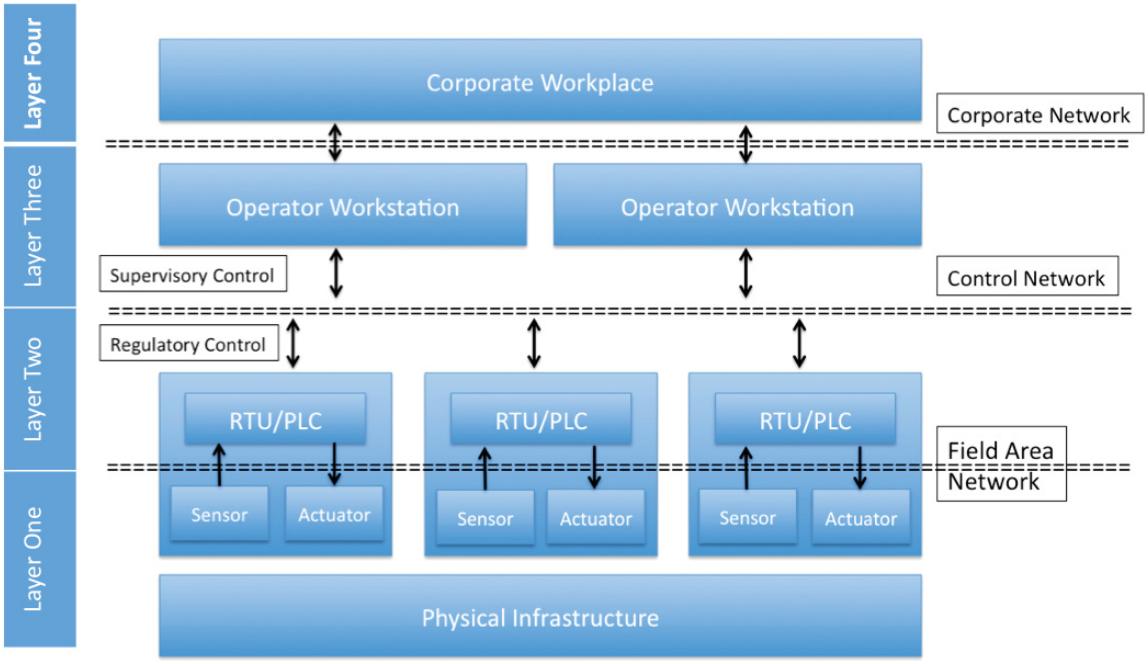


Figure 2.1: A typical control system architecture [8].

sub-system controlled by one or more controllers. Although the distinction of distributed control varies widely from the essence of SCADA systems, DCS basic components are similar to SCADA systems: HMI, master terminal unit (MTU), and field devices.

### *2.1.1 Master Terminal Unit.*

The MTU is the system controller, located in the control center. The MTU issues commands to field devices in remote locations primarily serving to gather, store, and process data [3]. Ultimately, all data is provided to the operator, via the HMI, as human readable information in the form of pictures and tables.

### *2.1.2 Human-Machine Interface.*

The HMI enables communication between the MTU and the human operator. Readable data transmitted from the MTU to the HMI is displayed graphically to the operator in the form of a mimic diagram. A mimic diagram provides a schematic representation of the remote processing location (Figure 2.2). The HMI, in conjunction

with the MTU, provides the operator access to RTUs and PLCs, thus allowing the operator to monitor and control remote processes.

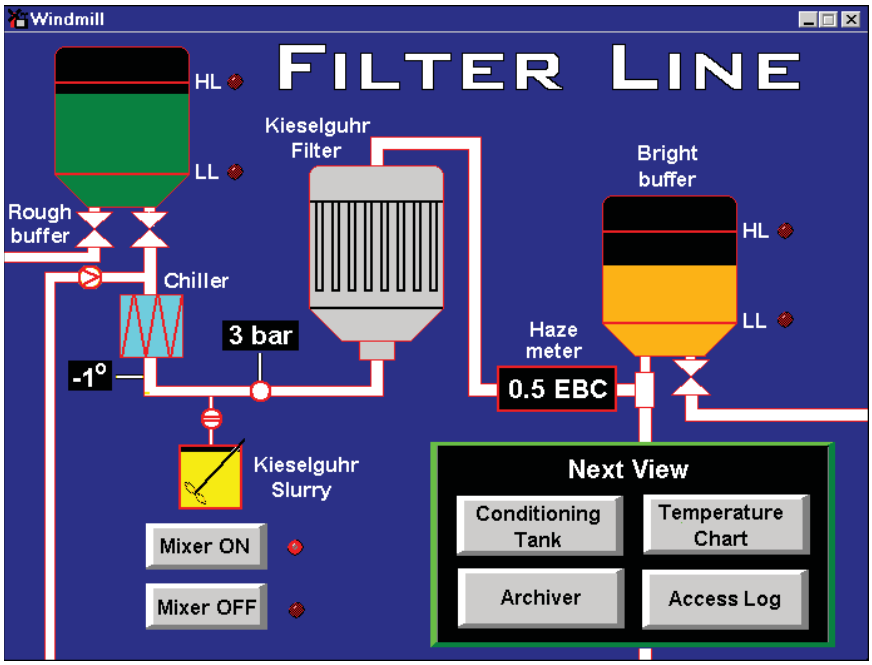


Figure 2.2: HMI mimic diagram [54].

### *2.1.3 Field Devices.*

RTUs are rugged industrial computers whose primary function is to interface with field devices, collecting telemetry data to transfer to the MTU. When an intelligent electronic device (IED) receives an instruction from the MTU, the RTU forwards the command (e.g., open or close a valve). Common types of IEDs include protective relaying devices, circuit breaker controllers, valves, and voltage regulators.

PLCs are specialized computers similar to RTUs. The distinguishing feature between PLCs and RTUs is the PLC's ability to conduct operation-based Boolean logic, thus providing the automation and regulatory control of industrial processes [3]. PLCs connect directly to field data interface devices, and incorporate programmed intelligence in the

form of logical procedures (e.g., Ladder Logic), which is executed in the event of certain field conditions [54]. Basic features of some PLCs include a web server and device specific communications protocols. The web server provides access to information from the control system using a web browser, while also allowing remote control system monitoring and modification. The device specific communications protocol service provides remote control and management of the programmable logic residing on the device.

#### *2.1.4 Industrial Control System Communications.*

ICS communication networks are comprised of the physical medium used to transfer data between the control center and field devices, and the device specific communication protocols. Typically ICSs employ one of three mediums: cable, telephone, or radio [3]. Since ICS inception, numerous proprietary protocols have emerged, but Modbus, distributed network protocol (DNP3), and Ethernet industrial protocol (EtherNet/IP) have emerged as the most prolific protocols [10]. The Modbus messaging protocol was developed in 1979 by Modicon to establish master-slave/client-server communication between intelligent devices, allowing for communication of up to 247 devices and later incorporated TCP [35]. DNP3 is a set of protocols developed for communications between various types of data acquisition and control equipment. In 2010, the DNP3 Technical Committee, in coordination with Institute of Electrical and Electronics Engineers (IEEE), established DNP3 as the standard for Electrical Power System Communications (i.e., IEEE 1815) [14]. EtherNet/IP, originally developed by Rockwell Automation in 2001, currently managed by the Open DeviceNet Vendors Association (ODVA), is an application layer protocol similar to simple network management protocol (SNMP) implementing the common industrial protocol (CIP) over transmission control protocol/Internet protocol (TCP/IP) [38].

## 2.2 ICS Security

### *2.2.1 Requirements.*

Within the United States, ICSs control the majority of key critical infrastructure to include power, water, transportation, and financial systems [49]. As such, the security of ICSs is integral to United States national defense. Concerns for the security of these systems has been expressed in multiple instances over the past two decades. In 1997, under Presidential Decision Directive 63 (PDD-63), President Bill Clinton created the Commission on Critical Infrastructure Protection to discuss the threat to control systems and the potential effects a successful attack could have on the electric power and oil and gas industries [31]. In 2001, after the attacks on the World Trade Center, the United States Congress enacted the USA Patriot Act (H.R. 3162), that included the Critical Infrastructure Protection Act of 2001, which states, "any physical or virtual disruption of the operation of the critical infrastructures of the United States be rare, brief, geographically limited in effect, manageable, and minimally detrimental to the economy, human and government services, and national security of the United States" [36].

In 2002, the National Research Council (NRC) identified the potential for attack on control systems as requiring urgent attention, finding that security experts reported 70% of energy and power companies experienced at least one severe cyber attack [19]. In 2003, President George W. Bush demonstrated concern regarding the threat of organized cyber attacks capable of causing debilitating disruption to national critical infrastructures, specifically noting the disruption could have significant consequences for public health and safety and emphasizing that the protection of control systems has become a national priority [19]. The National Institute of Standards and Technology (NIST) posits a cyber attack on energy production and distribution systems could endanger public health and safety, damage the environment, and have serious financial implications [19]. Economist Scott Borg projects that if a third of the country lost power for three months, the economic

price tag would be $700 billion, “a greater economic damage than any modern economy ever suffered...greater than the Great Depression...greater than the damage the United States did with strategic bombing on Germany in World War II” [33].

To understand the security concerns surrounding ICSs, it is first required to understand their inherent fragility. The real-time nature of ICSs requires system security to focus primarily on availability while also simultaneously introducing physical safety, performance, and graceful degradation [55]. By virtue of this prioritization, ICSs were designed to maximize performance, reliability, and efficiency, not designed for security. ICS security initially relied upon robust physical protection and network obscurity. A demand for increased availability, advances in technology, and the highly distributed nature of control systems has led to a demand for network connectivity, complexity, and extensibility introducing a new level of security threats and vulnerabilities. The call for connectivity led to a migration to TCP/IP as the predominant communications protocol suite used in connecting network hosts and the eventual widespread connection to the Internet. In 2011, Symantec assessed the threat to SCADA as critical, citing 129 public control system vulnerabilities, illustrating a substantial increase over the 15 vulnerabilities in 2010 [46]. ICS-CERT cited 171 unique vulnerabilities affecting ICS products as of December 2012 [50]. The expansion from physically separated closed networks to Internet connectivity exposed ICSs to not only specific device targeting, but traditional IT security attacks.

The evolution, from a closed network to Internet connectivity, not only exposed control systems to a vast number of threats and threat vectors, it also resulted in a convergence of control system security and traditional IT security [55]. IT security prioritization runs counter to ICS security, prioritizing confidentiality and integrity above availability. As ICSs are considered real-time operating systems and designed to operate for years without rebooting or interruption, traditional IT security practices are difficult to

perform due to the potentially disastrous effects on the core principles of ICS security: availability, reliability, performance, and safety [55]. Typical IT security and network administration practices include software updates, equipment upgrades, data encryption, anti-virus software, network assessments, and penetration testing. Andres Andreu, chief architect and vice president of engineering for Bayshore Networks, a leading ICS security firm, states, "There are a lot of controllers out there from the 1960's and 1970's that can't handle sophisticated security; PLCs with bytes of memory, unable to handle anymore information, let alone updates" [22]. Eric Byres, Chief Technology Officer (CTO) of Belden's Tofino Security, stated one vendor his firm works with estimates that less than 10 percent of its customers download the PLC patches it issues [22]. Data encryption and anti-virus software, two vital aspects of IT security, can cause network latency that negatively impacts overall performance [45]. A task such as mapping the network to identify hosts, operating systems, ports, and services can have catastrophic effects. In one example, upon performing a ping sweep of an active SCADA network controlling a 9-foot robotic arm, it was noted that one arm became active and swung around 180 degrees [45]. In another example, a ping sweep of an ICS network to identify hosts caused a system controlling the creation of integrated circuits in the fabrication plant to lockup, resulting in the destruction of $50,000 worth of wafers [45].

Penetration testing, a vital resource to network administrators, suffers from similar problems concerning ICS networks. Penetration testing involves simulating an attack from the perspective of a potential attacker, focusing on vulnerability discovery in order to strengthen network defense [42]. Because penetration testing involves active exploitation of security vulnerabilities, the potential risk to live production ICSs is high and the potential to crash a network consequently prevents penetration testing implementation. For example, in one incident, a natural gas utility hired an IT security consulting firm to test their corporate network. During the assessment, the consultants inadvertently

accessed the ICS network, causing the ICS to lock up, ultimately preventing the utility from sending gas through its pipelines for over four hours [45]. Based on these and similar incidents, ICS operators are hesitant to implement traditional IT security protocols. Consequently, the relationship between ICS security and management continues to clash with traditional network security practices, ultimately leaving ICS networks and devices vulnerable to exploitation.

United States Computer Emergency Response Team (US-CERT) highlights three areas wherein ICSs show the greatest vulnerability to attack: software security, configuration, and network security [49]. ICS software commonly suffers from the lack of secure software design and coding practices, leaving ICS network protocols and associated server applications prone to man-in-the-middle (MITM) data attacks, unvalidated user input, and subject to considerable information leakage through vulnerable custom ICS web services [49]. Many ICS individual component vulnerabilities are dependent on specific device implementation and include: un-patched operating systems, applications, and service vulnerabilities; failure to configure and implement applications and services securely (e.g., selecting security options and protecting credentials); default passwords; weak password policies; user accounts, applications, and services with administrator permissions; default security features; and open network connections [49].

US-CERT claims ICS networks are particularly susceptible to attacks due to the lack of defined security perimeters, network segmentation, and functional demilitarized zones (DMZ) [49]. In addition, firewalls meant to protect these networks do not take into account the traffic meant for ICSs and associated devices. Finally, it is a reoccurring theme throughout the ICS community that the network architecture is poorly understood, out-dated, and unsecured, which is compounded by weak enforcement of remote login policies and insufficient methods for monitoring and controlling network events [49].

### *2.2.2 Attacks.*

ICSs are the foundation of United States critical infrastructure and by their vary nature are prime targets for attack [33]. The DHS defines cyber threats to a control system as, "persons who attempt unauthorized access to a control system device and/or network using a data communications pathway, either trusted internal users or remote exploitation by persons unknown via the Internet" [49]. The Government Accountability Office (GAO) ICS threat table indicates general threats to control systems can originate from numerous sources; however, deliberate threats emanate from specific sources, including hostile governments, terrorist groups, disgruntled employees, industrial spies, organized crime, and hacktivists [45]. The motive for attack is varied, including competitive industrial advantage, information warefare, extortion, financial gain, revenge, and terrorism. Attack vectors include spear-phishing, default authentication implementation, direct Internet accessibility, and back-doors.

Figure 2.3 provides a timeline of notable control system attacks for the past two decades according to DHS and NIST. In 2000, an attack against the Maroochy Shire Sewage facilities Queensland, Australia resulted in over 200,000 gallons of raw sewage spilling into local parks and rivers severely impacting marine ecosystems and the local economy [1]. Vitek Boden, a former employee of Hunter Watertech, an Australian firm specializing in the installation of SCADA radio-controlled sewage equipment, perpetrated this attack in retaliation for the Maroochy Shire Council declining his bid for a job. Boden decided to get revenge on both the Council and his former employer by packing his car with stolen radio equipment and driving around the area on at least 46 occasions from February 28 to April 23, 2000, issuing radio commands to the sewage equipment [1]. His actions went unabated for over two months until a traffic violation following an attack caused him to fall under suspicion and ultimately resulted in his arrest. This incident

became the first widely known control system attack, exposing the real-world impact while also demonstrating how difficult it is to catch an attacker.

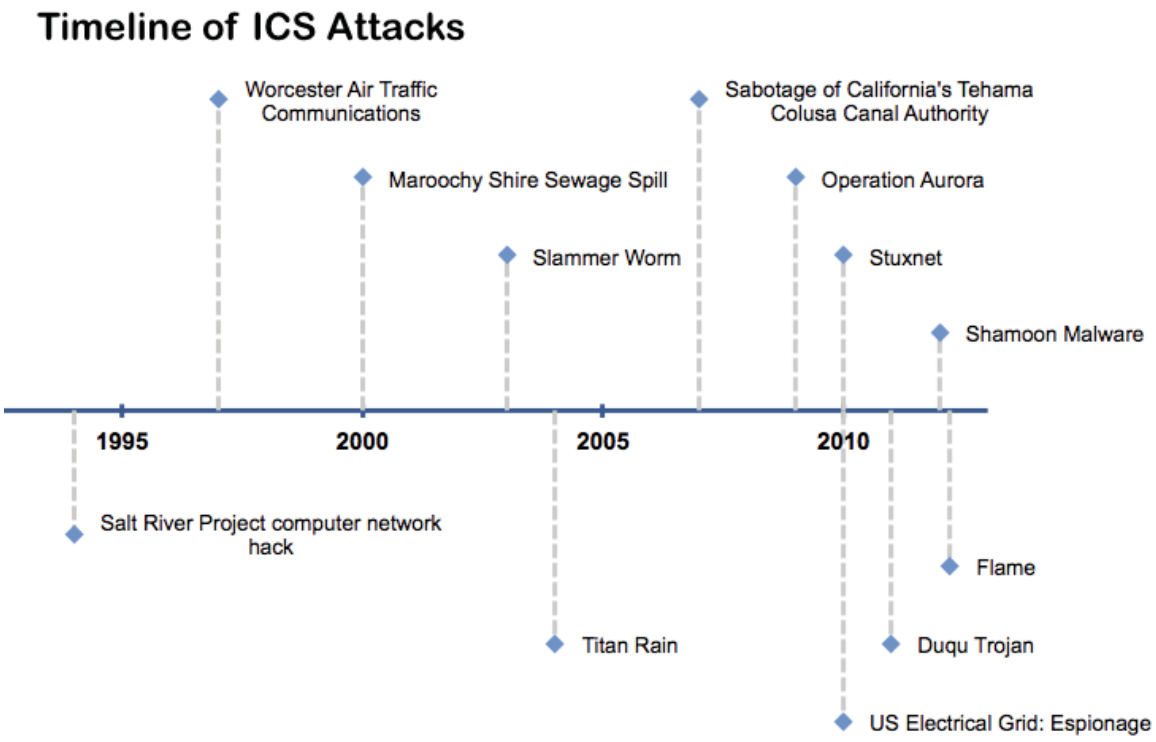


Figure 2.3: Timeline for notable control system network attacks [27].

In 2003, the Davis-Besse nuclear power plant in Oak Harbour, Ohio, was infected with the Microsoft SQL Slammer worm [47]. The worm infiltrated a private computer network and caused a network traffic overload on the site. As a result, the Safety Parameter Display System (SPDS) was inaccessible for almost 5 hours, and the plant process computer was inaccessible for over 6 hours [47]. The SPDS, a vital component of the plant Emergency Response Facility Data System (ERFDS), monitors physical plant parameters to include temperature, pressure, level, valve position, radiation level, and flow. The worm also disrupted communications on the control networks of at least five other utilities by propagating so quickly that control system traffic was blocked [47].

Forensic analysis determined a rogue T1 connection bypassed the firewall and access control policies, allowing access to the control network. The worm originated from a software consulting firm's infected server, utilizing the aforementioned T1 connection. Fortunately, the plant was idle during the time of the attack; hence no significant safety issues occurred.

In 2007, the DHS launched project Aurora designed to demonstrate a cyber attack against a generator [41]. The experiment involved infiltrating a replica power plant control system and changing the operating cycle of the generator, ultimately resulting in generator shutdown. Project Aurora proved empirically, in a research setting, the ability to attack a physical device via the Internet.

"Stuxnet was a game-changer because it opened people's eyes to the fact that a cyber event can actually result in physical damage," says Mark Weatherford, deputy undersecretary for cyber-security in the National Protection Programs Directorate at the U.S. Department of Homeland Security [34]. In June 2010, the Belorussian security VirusBlokAda first discovered Stuxnet, a randomly propagating worm with payloads targeting specific ICS [34]. Stuxnet spread using traditional security vulnerabilities in commercial operating systems, and then propagated to two ICS applications (with hard-coded passwords) to inhibit the functioning of Variable Frequency Drives (VFD) made by specific vendors [30]. To remain hidden, the worm displayed the last program sent to the VFDs while running its own code, similar to running a closed circuit television (CCTV) in a loop. The worm was designed to increase and decrease centrifuge speeds causing the aluminum housing to expand and contract, ultimately coming into contact with other centrifuges.

#### *2.2.3 Trending.*

ICS Internet connectivity is based on a demand for increased availability, advances in technology, and the highly distributed nature of control systems. The migration to TCP/IP

as the predominant communications protocol suite solidified widespread ICS Internet connectivity. In 2007, the British Columbia Institute of Technology conducted a study of 47 control system cyber incidents occurring between 2002 and 2006 which reported a remote point of entry as the threat vector [7]. The results indicated that while the business network was a major source, secondary pathways such as dial-up connections, wireless systems, public telecommunications networks, VPNs, and third-party connections were all significant contributors [7]. Figure 2.4 details the results of the study indicating direct Internet accessibility as tied for the third most often utilized attack vector.

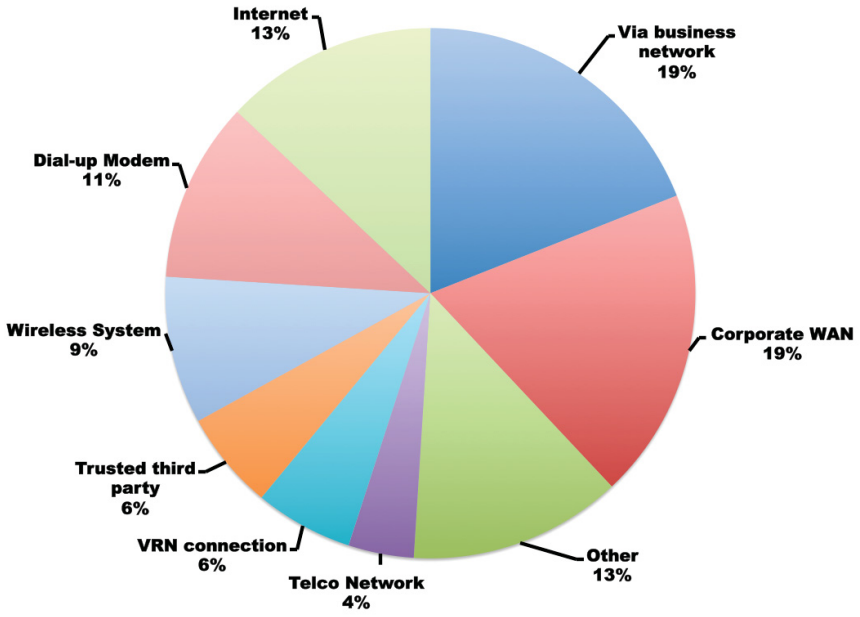


Figure 2.4: Critical infrastructure cyber threat vectors - remote points of entry [7].

In 2008, during a SANS SCADA Security Conference in New Orleans, Tom Donahue, Central Intelligence Agency (CIA) senior analyst, confirmed the remote cyber exploitation of electricity utilities outside the United States [9]. Donahue presented a written statement reading:

> We have information, from multiple regions outside the United States, of cyber intrusions into utilities, followed by extortion demands. We suspect, but cannot confirm, that some of these attackers had the benefit of inside knowledge. We have information that cyberattacks have been used to disrupt power equipment in several regions outside the United States. In at least one case, the disruption caused a power outage affecting multiple cities. We do not know who executed these attacks or why, but all involved intrusions through the Internet [9].

In 2009, Idaho National Laboratory (INL) conducted a study forecasting the cyber threat to critical infrastructure from 2010 to 2015, specifically trending control system exposure. The results of this study forecasted a proliferation of control systems, increased digital and IP base, expanded use of wireless communications, and lagging security measures [17]. Specifically, INL predicts the world ICS market to grow at a 8.9% rate into 2015 [17]. Figure 2.5 approximates PLC market growth worldwide through 2015, resulting is a $17 trillion market.

In addition to ICS proliferation and increased connectivity, ICS is gaining notoriety across the whitehat and blackhat communities, typically for their vulnerability and exposure. Security conferences such as DEFCON, Blackhat, RSA, and SANS routinely provide presentations on ICSs to include attack vectors and specific device exploitation [8].

## 2.3 Shodan

In 2009, programmer John Matherly launched Shodan, a computer search engine supplying a graphical user interface capable of readily identifying Internet-facing devices [43]. More specifically, Shodan identifies any device with a routable IP address to include computers, printers, web-cams, and ICS devices. Shodan crawls the Internet indexing devices and interrogating available services. The bulk of the data retrieved by Shodan is

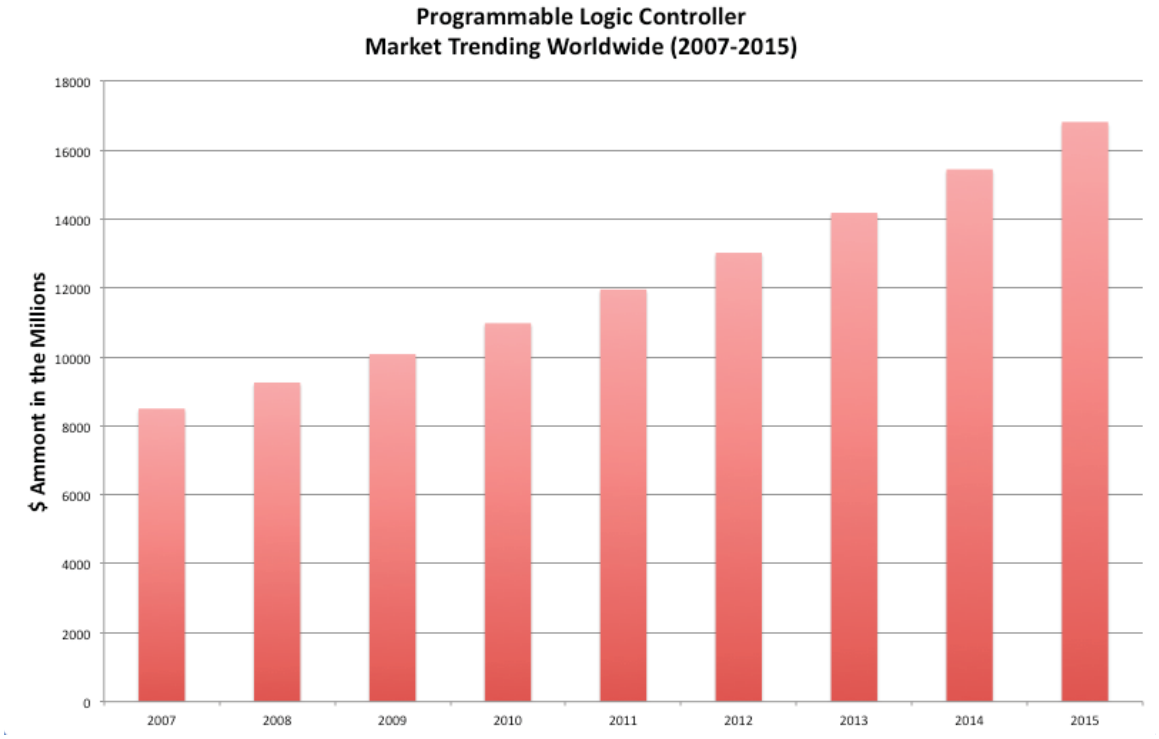


Figure 2.5: INL PLC market trending indicating ICS proliferation [17].

taken from “banners”, which are comprised of meta-data a server sends back to the client [43]. Note that the port 80 service banners obtained by Shodan are in reality HTTP headers and for consistency this research will use the term banner as indicated in Shodan documentation. Shodan stores the device IP address, ports, and service banner data in a searchable database accessible via the Shodanhq.com web interface or through the Shodan application programming interface (API). Users are able to query the Shodan database using a series of filters to include: country, hostname, net (i.e., specific IP range), operating system (OS), and port.

Initially, Shodan began by interrogating basic ports to include port 21 (i.e., FTP), 22 (i.e., SSH), 23 (i.e., Telnet), and 80 (i.e., HTTP), but has since widely expanded port interrogation to 40 services (Table 2.1) [43]. In addition to an Internet-facing device

index, Shodan offers an exploit database, a raw nmap data output visualization tool, and an enumeration module built into the metasploit exploitation framework.

Table 2.1: Shodan documented service interrogation filters [43].

| Port | Service | Port | Service |
|---|---|---|---|
| 21 | FTP | 1900 | UPnP |
| 22 | SSH | 2323 | Telnet |
| 23 | Telnet | 3306 | MySQL |
| 25 | SMTP | 3389 | RDP |
| 53 | DNS | 5000 | Synology |
| 80 | HTTP | 5001 | Synology |
| 81 | HTTP | 5432 | PostgreSQL |
| 110 | POP3 | 5560 | Oracle |
| 119 | NNTP | 5632 | PC Anywhere |
| 137 | NetBIOS | 5900 | VNC |
| 143 | IMAP | 6379 | Redis |
| 161 | SNMP | 7777 | Oracle |
| 443 | HTTPS | 8000 | Qconn |
| 445 | SMB | 8080 | HTTP |
| 465 | SMTP | 8129 | Snapstream |
| 623 | IPMI | 8443 | HTTPS |
| 993 | IMAP+SSL | 9200 | ElasticSearch |
| 995 | POP3+SSL | 11211 | MemCache |
| 1023 | Telnet | 27017 | MongoDB |
| 1434 | MS-SQL | 28017 | MongoDB Web |

In October 2010, ICS-CERT published a Control Systems Analysis Report (CSAR-10-025-01 Analysis of Shodan Computer Search Engine) detailing the Shodan search engine's capability of identifying potentially vulnerable control system interfaces, as well as discussing the importance of minimizing network exposure by ensuring that control system devices are not visible on the Internet [48]. Subsequently, ICS-CERT released five ICS alerts (ICS-ALERT-10-301- 01, ICS-ALERT-10-301-01A, ICS-ALERT-11-343-01A, ICS-ALERT-12-046-01, ICS-ALERT-12-046-01A), detailing further concerns over Shodan's ability to identify Internet-facing ICS devices [12].

In 2011, Eireann Leverett used Shodan to counter claims of ICS network segregation. Leverett presented two years of historical evidence, providing timelines and geo-location of over 7,500 ICS devices connected to the Internet to include: HVAC systems, building management systems, meters, HMIs, and PLCs [30]. Leverett used 29 specific Shodan search queries to identify ICS devices. Table 2.2 provides a comparison between Leverett's 2011 query results and the same queries executed in 2013 in support of this research. In approximately two years, the number of identified devices dramatically increased from 7,500 to 57,409.

Leverett's research highlights Shodan's ability to identify global ICS exposure, providing a reconnaissance tool for attackers. Leverett asserts, "databases of vulnerable critical national infrastructure will be traded in the future like the data of stolen credit card numbers today; and as such, the ability to rapidly act in an automated manner on such data by either defenders or attackers will define the next few years of critical infrastructure protection" [30].

In 2012, Bob Radvanovsky and Jake Brodsky of InfraCritical launched Project SHINE (Shodan Intelligence Extraction) to counter claims of ICS network segregation and expose Internet-facing devices [6]. Project SHINE used the Shodan API and over 700 specifically designed queries to identify vulnerable Internet-facing ICS devices. Project

Table 2.2: Leverett's Shodan search results - 2011 vs 2013.

| **Shodan Query** | **2011** | **2013** | **Category** | **Inc/Dec** |
|---|---|---|---|---|
| A850+Telemetry+Gateway | 3 | 34 | Telemetry | 1033% |
| ABB+Webmodule | 3 | 3 | Embedded Webserver | 0% |
| Allen-Bradley | 23 | 99 | PAC | 2533% |
| /BroadWeb/ | 148 | 352 | HMI | 6800% |
| Cimetrics+Eplus+Web+Server | 6 | 16 | Embedded Web Server | 333% |
| CIMPLICITY | 90 | 239 | HMI | 4967% |
| CitectSCADA | 3 | 3 | PCS | 0% |
| EIG+Embedded+Web+Server | 104 | 137 | Embeddded Web Server | 1100% |
| eiPortal | 1 | 98 | Historian | 3233% |
| EnergyICT | 585 | 2706 | RTU | 70700% |
| HMS+AnyBus-S+WebServer | 40 | 121 | Embedded Web Server | 2700% |
| i.LON | 1342 | 4643 | BMS | 110033% |
| ioLogik | 36 | 184 | PLC | 4933% |
| Modbus+Bridge | 12 | 99 | Protocol Bridge | 2900% |
| ModbusGW | 11 | 94 | Protocol Bridge | 2767% |
| Modicon+M340+CPU | 3 | 56 | Protocol Bridge | 1767% |
| Niagara+Web+Server | 2794 | 34560 | HAN/BMS | 1058867% |
| NovaTech+HTTPD | 1 | 0 | Embedded Web Server | -33% |
| Powerlink | 257 | 3121 | BMS/HAN | 95467% |
| Reliance+4+Control+Server | 10 | 6 | SCADA | -133% |
| RTS+Scada | 15 | 28 | SCADA | 433% |
| RTU560 | 2 | 18 | RTU | 533% |
| Simatic+HMI | 9 | 91 | HMI | 2733% |
| SIMATIC+NET | 13 | 152 | HMI | 4633% |
| Simatic+S7 | 13 | 201 | PLC | 6267% |
| SoftPLC | 80 | 1088 | PAC | 33600% |
| TAC/Xenta | 1880 | 9165 | BMS | 242833% |
| WAGO | 2 | 89 | Telemetry | 2900% |
| webSCADA-Modbus | 3 | 6 | HAN | 100% |
| Total | 7489 | 57409 | | |

SHINE partnered with DHS to identify over 500,000 Internet-facing ICS related nodes worldwide [6]. Further coordination with ICS experts and ICS-CERT narrowed the results to 7,200 devices, with many lacking even the most basic security precautions and using weak, default, or no authentication [50].

In 2013, a researcher presented his findings on Internet-exposed critical infrastructure devices at the SANS North American ICS and SCADA Summit [5]. The researcher demonstrated two examples for identifying Internet-facing ICS devices by using Google search queries. Indeed, the researcher was able to identify Siemens S7 PLCs and Allen-Bradley ControlLogix PLCs exposed to the Internet. In one case, the researcher was able to use readily available Internet tools to identify the water facility being controlled and affirm operating characteristics, physical location, and the individual responsible for the system. Table 2.3 and Table 2.4 outline the security concerns and implications for each device based on direct Internet accessibility. This research reaffirms the inherent insecurity of ICSs and exemplifies the potential for Internet-facing ICS device exploitation.

Table 2.3: Siemens S7 - Security concerns and implications [5].

| | |
|---|---|
| Security Concerns | 1. Listening on port 102 allows device management over TCP<br>2. Listening on port 21 and port 80<br>3. Firmware version outdated<br>4. Able to map network architecture |
| Implications | 1. SSA-724606: It is possible to cause device to go into defect mode by sending specially crafted packets to port 102<br>2. ICS-ALERT-11-204-01B: Undocumented functions allow access to internal diagnostics via undocumented password |

Table 2.4: Allen-Bradley ControlLogix 1769 - Security concerns and implications [5].

| | |
|---|---|
| Security Concerns | 1. Listening on diagnostic port 44818<br>2. Firmware version outdated<br>3. Configured for Remote Run<br>4. Security set to No Protection |
| Implications | 1. Complete control of device and operating parameters<br>2. Ability to connect and retrieve project file<br>3. Ability to push new project file and firmware<br>4. Ability to disable PLC access to internal diagnostics via undocumented password |

## 2.4 Related Research

### *2.4.1 Honeypots.*

A honeypot is an information system resource whose value lies in unauthorized or illicit use of that resource [44]. Indeed, a honeypot is a decoy server or system designed to gather information regarding an attack or intrusion into a network or system [15]. Honeypots are categorized as either low-interaction or high-interaction. Low-interaction honeypots offer limited interaction, utilizing service and operating system emulation, and provide an easily deployable security mechanism with minimal risk [44]. Table 2.5 details the advantages and disadvantages of low-interaction honeypots.

High-interaction honeypots are typically more complex and designed to imitate the activities of real systems by hosting a variety of services through the use of real operating systems and applications [44]. High-interaction honeypots appear more realistic and appealing from an attackers perspective, but also increases the risk of the honeypot as attackers may use the real operating systems to attack non-honeypot systems. Table 2.6 details the advantages and disadvantages of high-interaction honeypots.

Table 2.5: Low-interaction honeypots: Advantages vs Disadvantages [44].

| **Low-interaction**: Emulated operating systems and services | |
|---|---|
| Advantages | • Easy to install and deploy. Typically requires installing and configuring software on a computer.<br>• Minimal risk, as the emulated services control what attackers can and cannot do. |
| Disadvantages | • Captures limited amounts of information, mainly transactional data and some limited interaction. |

Table 2.6: High-interaction honeypots: Advantages vs Disadvantages [44].

| **High-interaction**: No emulation, real operating systems and services are provided | |
|---|---|
| Advantages | • Able to capture far more information, including new tools, communications, or attacker keystrokes.<br>• No assumptions on how an attacker will behave; provides an open environment capturing all activity. |
| Disadvantages | • Can be complex to install or deploy (commercial versions tend to be much simpler).<br>• Increased risk, as attackers are provided real operating systems to interact with. |

Honeypots are further characterized according to security goals: prevention, detection, reaction and research [44]. Prevention honeypots are designed to stop the attacker from attacking the production system by employing IP address deception, network attack simulations, and information deception. Detection honeypots provide real-time alerts based on system activity rather than known signatures as with a typical IDS. Reaction honeypots accompany a production system and mirror their setup with the primary function to identify exploitations and patch vulnerabilities. Research honeypots are designed to invite malicious attacks by incorporating common vulnerabilities and OS security holes.

### *2.4.2 Honeynet Project.*

The Honeynet project began in 1999 as a research activity to evaluate and explore the use of honeypots and honeynets to increase the knowledge of attackers' behaviors, motivations, attack tools, and other relevant vulnerability data [28]. In 2004, the honeynet project was expanded to include ICS honeypots with the primary goal of determining the feasibility of designing a software framework to simulate a variety of industrial networks and devices [16]. Necessary requirements for the framework included:

- A targetable platform capable of allowing users to gather data on attacker trends and tools.
- A scriptable industrial protocol simulator to test a live protocol implementation.
- Research countermeasures, such as device hardening, stack obfuscation, reducing application information, and the effectiveness network access controls.

In 2005, the project culminated in a PLC honeynet developed by Venkat Pothamsetty and Matthew Franz from the Critical Infrastructure Assurance Group (CIAG). The PLC honeynet utilized Honeyd to simulate standard PLC services: TCP/IP stack of an Ethernet-based device, Modbus services, SNMP, Telnet, file transfer protocol (FTP), and hypertext transfer protocol (HTTP) [16]. Honeyd is a low-interaction honeypot developed

by Niels Provos capable of simulating a virtual computer system at the network level [44]. Perhaps most importantly, Honeyd is capable of implementing python scripts simulating basic PLC device service interaction. The CIAG PLC honeynet offers a basic framework for ICS honeypots, but the reliance on Honeyd limits overall device interaction. In addition, Honeyd's device service scripting utilizes a python HTML implementation which inhibits the ability to mimic service such as the Allen-Bradley ControlLogix web server. While the CIAG PLC honeynet provided a basis for future ICS honeypot research, the project is no longer maintained.

#### *2.4.3 Digital Bond SCADA Honeynet.*

In 2006, Digital Bond created a virtual PLC honeynet providing both a monitoring system and a simulated PLC target device, designed to aid researchers in understanding the potential risks of exposed control system devices [13]. Figure 2.6 details the virtual honeynet consisting of two virtual machines (VMs), a Target VM and a Honeywall VM, running on a single host using VMware. All incoming and outgoing traffic is captured by the Honeywall acting as a transparent bridge. Management is designed for remote access directly to Honeywall for reports and raw packet captures.

The Generation III Honeywall includes Snort IDS in packet capture mode, Digital Bond SCADA IDS signatures, Sebek, Argus, Walleye, and MySQL [13]. Honeywall report generation includes top ten scanned ports, the top ten remote IPs, the number of packets in and out of the network, and the total number of Snort alerts generated. The target ICS device uses Honeyd to simulate a Schneider Modicon Quantum PLC and expose several basic services to include: Modbus TCP, Telnet, FTP, HTTP, and SNMP. Table 2.5 outlines the available services provided by the target system as well as a brief description of their purpose.

Later updates to the Digital Bond honeynet included the ability to utilize a physical ICS device rather than the simulated target system.

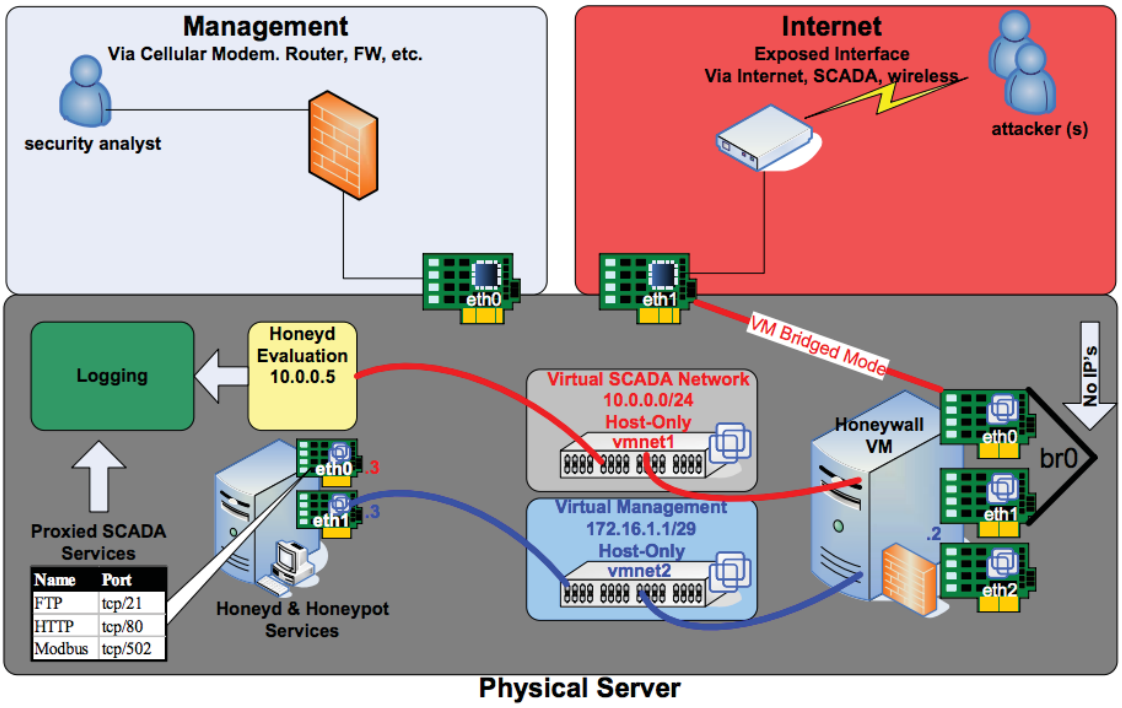


Figure 2.6: Digital Bond SCADA honeynet architecture [13].

Table 2.7: Digital Bond ICS honeynet target system services [13].

| Service | Port | Purpose |
|---|---|---|
| FTP | tcp/21 | Firmware/Device Management |
| Telnet | tcp/23 | Device Configuration/Management |
| HTTP | tcp/80 | Device Configuration/Management |
| SNMP | udp/161 | Device/Service Health/Statistics |
| Modbus TCP | tcp/502 | Monitoring and Control |
| VxWorks Debugger | tcp/17185 | Device Debugger |

### *2.4.4 The Honeynet Project - Conpot.*

In 2013, the Honeynet Project released their first ICS honeypot, Conpot, supporting Modbus and SNMP protocols [25]. The default configuration of Conpot simulates a basic Siemens SIMATIC S7-200 PLC. This project has been deployed worldwide in an effort to identify attack vectors for ICS devices. One such long-term deployment from the United Kingdom Honeynet Project Chapter includes 43 low interaction sensors, resulting in over 2,000,000 attacks and 36,000 attacker IPs in 2012 alone [25]. Note that this open-source venture relies on a community of volunteer security experts and lacks reliable source for support, to include basic tool upgrades to match recent operating system standards.

### *2.4.5 Iowa State University.*

In 2011, an Iowa State University student deployed the Digital Bond SCADA honeynet to measure specific ICS device targeting in a post-Stuxnet world [51]. The Digital Bond honeynet was deployed for 38 days using a single server to host the Digital Bond Honeywall and simulated Modicon PLC. Data collection included raw packet captures and intrusion detection reports generated by the Digital Bond Honeywall. Figure 2.7 details the honeynet architecture. The primary goal was to identify SCADA PLC specific targeting. Of particular interest was any interaction with the Modbus or VxWorks Debugger services, TCP port 502 and user datagram protocol (UDP) port 17185.

The findings were categorized into SCADA specific attacks and traditional IT attacks. While the research did not identify any instances of SCADA specific targeting, numerous traditional IT attacks were identified targeting the PLC. Figure 2.8 outlines the details of traditional IT targeting of the Modicon PLC.

This research was primarily limited by the deployment location. The honeypot was deployed internal to the Iowa State University and behind university network defenses. As a result, the majority of device interaction is assumed to have occurred by university

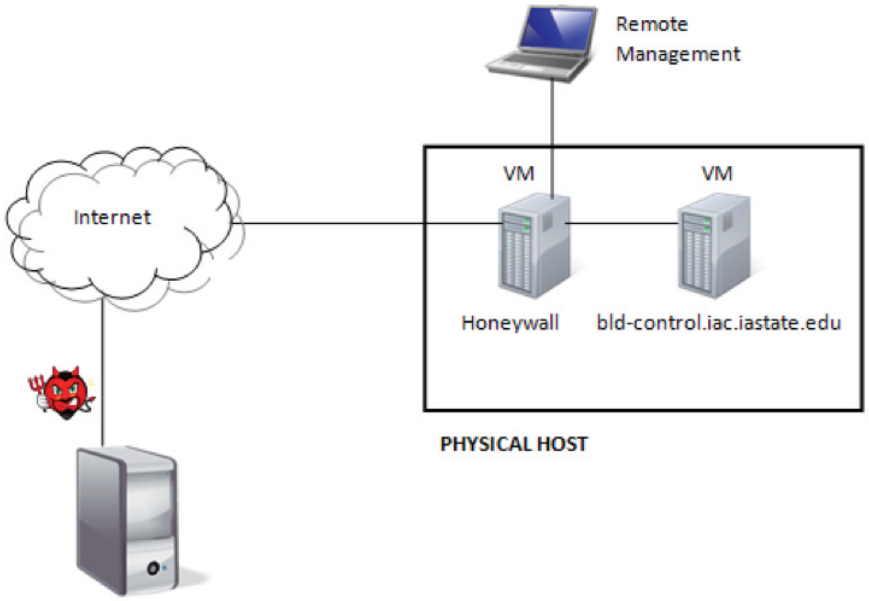


Figure 2.7: Iowa State University Digital Bond SCADA honeynet deployment architecture [51].

students or staff. It was also noted the deployment network functions as the university cyber attack educational testing network, which may account for the type and level of Snort alerts and as well as limited ICS specific device interaction.

### *2.4.6 Trendmicro.*

In 2013, Kyle Wilhoit, Trendmicro researcher, published a series of reports detailing his efforts to expose malicious targeting of Internet-facing ICSs [29]. Wilhoit's first honeynet deployment lasted 28 days and included three honeypots deployed in geographically-separated locations throughout the United States. Table 2.8 provides details of each honeypot. The first honeypot provided a simulated water pressure station via Honeyd implementation. The second honeypot simulated a HMI via a physical server running PLC software. The third honeypot simulated a factory temperature control system via a physical PLC. Wilhotit defined an attack as anything deemed a threat to

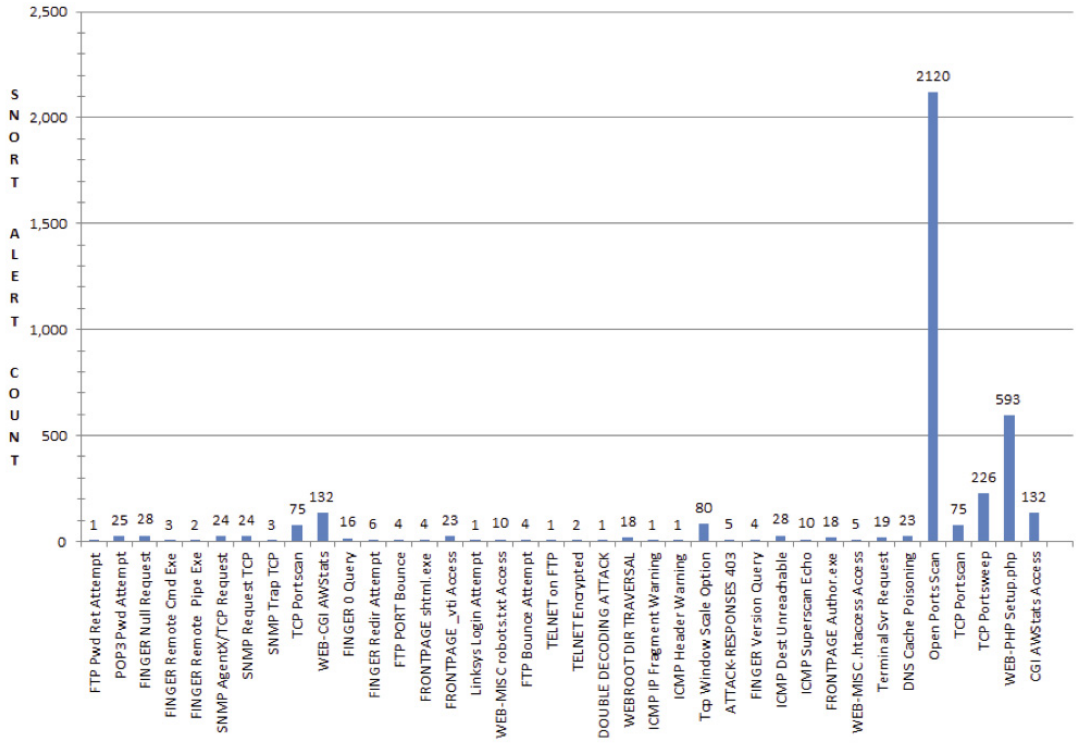


Figure 2.8: Research results indicating traditional IT attacks indiscriminately targeting PLC honeypot [51].

Internet-facing ICS devices to include unauthorized access to secure areas of sites, modifications on perceived controllers, or any attack against a protocol specific to ICS devices [52].

Wilhoit's deployment included tactics for optimizing the sites for searches, publishing the sites on Google, utilizing specific server naming conventions (e.g., SCADA-1, SCADA-2), and seeding the devices within Shodan [53]. Wilhoit's findings revealed all honeypots were attacked within 18 hours of deployment and within the first month the honeypots registered 39 attacks from 14 different countries, 12 of which were classified as targeted attacks [53]. Attacks included: attempted Modbus traffic modification, attempted access to secured PHP pages, attempted malware exploitation,

Table 2.8: Wilhoit honeynet deployment: honeypot design [52].

| | |
|---|---|
| Emulated ICS honeypot | 1. Honeypot: Simulated water pressure station, high-interaction<br>2. Device: Emulated PLC via Honeyd<br>3. Location: Virtual instance of Ubuntu Amazon EC2<br>4. Services: HTTP (Apache web server), Modbus, FTP |
| Emulated ICS honeypot | 1. Honeypot: HMI, high-interaction<br>2. Device: Dell DL360 server running PLC software<br>3. Location: Physically deployed within the US<br>4. Services: HTTP |
| Physical PLC honeypot | 1. Honeypot: Factory temperature control system, high-interaction<br>2. Device: Triangle Research Nano-10 PLC<br>3. Location: Physically deployed within the US<br>4. Services: HTTP, Modbus |

and attempted device settings manipulation. Each attack was preceded by a port scan, following traditional network attack methodology [29]. Figure 2.9 depicts Wilhoit's breakdown of activity by country. In addition to attacks specifically targeting SCADA devices, Wilhoit identified an attempt to spearphish the site administrator of one of the honeypot devices.

In August 2013, Wilhoit's second honeypot deployment provided evidence of a malicious actor breaking into a simulated United States water control systems [29]. Wilhoit used a tool called the Browser Exploitation Framework, or BeEF, to gain access to the attackers' systems to triangulate their location using the built-in Wi-Fi cards [29]. Wilhoit's research extended the research of Leverett and Project SHINE, providing evidence of Internet-facing ICS targeting.

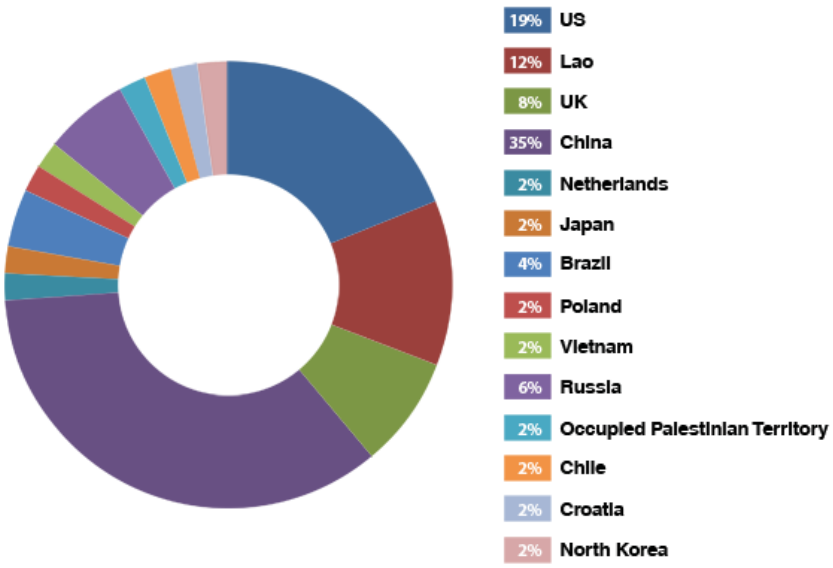


Figure 2.9: Wilhoit honeypot targeting breakdown by country [53].

## 2.5 Knowledge Gaps

A threat can be defined as a marriage of capability, intent and opportunity [17]. ICS devices are currently deployed Internet-facing [6, 30], Shodan is able to identify these devices, and evidence shows Internet-facing ICSs are being attacked [52]. By all indications Shodan should be categorized as a threat to Internet-facing ICSs, however, there is a lack of empirical evidence linking Internet-facing device targeting as a result of Shodan. More precisely, what is the utilization of Shodan as a reconnaissance tool for ICS device targeting?

## 2.6 Summary

United States critical infrastructure security is defined by the underlying ICSs security and the device exposure. ICSs are inherently fragile due to legacy equipment, inability to conduct traditional network security, and contrasting security priorities between ICS security and traditional IT security professionals. ICS trending shows a

proliferation of ICS and a dramatic shift to increase device availability and connectivity, while sacrificing overall device security. The demand for connectivity has extended beyond internal networking and direct access via corporate network to direct Internet-facing deployment. Research using the Shodan computer search engine identified thousands of ICS devices directly Internet-facing. Shodan, while not specifically designed to identify ICS devices, provides a passive reconnaissance tool capable of identifying ICS devices down to the specific make and model. Additional research provided evidence of malicious ICS exploitation, wherein actors attempted to exploit an ICS honeypot simulating a United States water utility. The threat of ICS exploitation is real; devices are deployed Internet-facing with default or weak authentication and Shodan is capable of identifying these devices. As such, this research attempts to fill the intelligence gap by evaluating the impact of Shodan on Internet-facing ICS device security.

# III. Methodology

This chapter discusses the goals, approach, design and implementation of the honeypots used to evaluate the impact of the Shodan computer search engine on Internet-facing ICS devices.

## 3.1 Problem Definition

Strategically, the intent of this research is to evaluate Shodan's impact on Internet-facing ICS device security by deploying a series of unsolicited, high-interaction ICS honeypots. The primary goals are to evaluate Shodan indexing functionality, contrast network activity levels as a result of Shodan identification, and enumerate any ICS specific targeting and attacks. Shodan indexing functionality is evaluated by determining Shodan's scanning routine, scanning frequency, and web database identification timeliness. Network activity is analyzed by measuring TCP connections, total TCP packet count, and the number of unique IP addresses interacting with each device. ICS specific targeting is evaluated by visual packet inspection and Snort IDS analysis with known ICS signatures. Visual packet inspection will identify device reconnaissance and unauthorized access to secure areas of the web management console. It is hypothesized Shodan will scan and index all devices within 30 days and further hypothesized post identification, each device will receive an increase in network activity, to include targeted ICS attacks.

The foundation of the Shodan computer search engine is a searchable database containing service banners for Internet accessible devices. Therefore, the secondary goal is to assess the impact of ICS device service banner data relative to device identification within Shodan and subsequently evaluate the ability to limit Shodan device exposure via banner manipulation. The service banner's impact is evaluated by comparing the number of targeted attacks as a result of Shodan identification of two honeypots, one presenting an

enticing service banner directly identifying the device make and model, while the second honeypot alters the service banner to obfuscate the device. It is hypothesized the honeypot presenting a more enticing service banner will see an increase in targeted attacks post Shodan identification as compared to both the standard and Obfuscated honeypots.

This research extends previous efforts to evaluate Internet-facing ICS device vulnerability by measuring the impact of the Shodan computer search engine. Specifically, this research presents an in-depth understanding of Shodan indexing functionality, develops a timeline for scanning, reveals the impact of Shodan identification on device network activity levels, and enumerates ICS specific targeting. In addition, this research evaluates the ability to limit device exposure by service banner manipulation. This data is critical to security professionals when crafting a defense strategy for ICS security.

## 3.2 Approach

To achieve the aforementioned goals, a honeynet comprised of four high-interaction honeypots is deployed unsolicited, Internet-facing, and co-located with an ICS integrator network. Each honeypot is representative of devices currently identifiable via Shodan. In addition, the devices are configured with default authentication settings to simulate newly deployed PLC devices and attract the broadest level of interaction. The honeynet is comprised of standard PLCs and banner mangled PLCs. The standard PLCs represent a baseline comparable to devices currently identifiable via Shodan, allowing the research results to indicate potential patterns across a larger population. The banner mangled honeypots measure the security implications of service banner data revealed by ICS devices. Data analysis and evaluation expands the understanding of Shodan's scanning functionality, compares network activity post Shodan identification, details ICS device targeting as a result of Shodan identification, and assesses the ability to limit device exposure via banner manipulation.

## 3.3 Motivation

ICS security is vital to United States national security due to their role in monitoring and controlling a majority of critical infrastructure systems supporting oil and gas pipelines, water distribution systems, electrical power grids, nuclear plants, and other manufacturing operations. Stuxnet [34] and David-Besse [47] offer historical examples of the ramifications of ICS exploitation. Stuxnet was a sophisticated, targeted attack infiltrating the deepest depths of the network, while David-Besse depicts the effects of a traditional network worm on an ICS network. The work of Leverett [30] and Project SHINE [6] demonstrate that critical infrastructure ICS devices are Internet-facing and readily identifiable via the Shodan computer search engine. Understanding the implications of targeted attacks against United States critical infrastructure ICSs, this research presents an investigation into the impact of Shodan on ICS security, specifically measuring network activity of Internet-facing devices as a result of Shodan identification. In addition, this research evaluates the ability to limit Shodan device exposure via service banner manipulation.

## 3.4 Setup and Deployment

Deploying a honeypot requires methodical planning, preparation, and understanding of the motivations for deployment [44]. Seven key decisions are required when planning honeypot deployment: location, deployment length, honeypot type, design configuration, setup validation, data collection, and evaluation.

### *3.4.1 Location.*

Honeynet deployment is highly dependent on available resources and research intent. This research uses high-interaction honeypots (i.e., physical devices) which precludes any virtual deployment locations such as the Amazon cloud. Locations considered for deployment include co-location with device manufacturers, ICS integrators, live ICS production enclaves, commercial business class IP space, and residential IP space. In

order to represent the most realistic honeypot it serves to deploy the honeypots co-located in a venue in some form associated with ICSs, removing the options of commercial or residential IP space. This research seeks to evaluate Shodan's impact on live production ICSs controlling critical infrastructure, therefore deployment co-located with device manufacturers is excluded. The resulting options include ICS integrators and live production environments. The honeynet is not co-located with live production environments primarily because of the inability to guarantee the devices would not interfere with the surrounding ICS environment. Additionally, the available production environment network architecture prevents the devices from direct Internet accessibility. Finally, this research is designed to attract the broadest swath of device interaction and, as such, would potentially draw unwanted malicious activity to the real-world ICSs, an unacceptable risk to live United States critical infrastructure.

The final option for deployment is co-location with an ICS integrator. ICS integrators specialize in bringing together multiple facets of ICS component subsystems into a single functioning system. This includes vendor coordination, system assembly, installation, maintenance, and security [11]. This location offers a unique opportunity to evaluate device targeting in a location prior to live ICS deployment. In the broader scope of ICS targeting, the ICS integrator offers a prime location for exploitation as they represent the middleman between the manufacturer and live production environment, a single choke point for the exploitation of multiple ICS venues. In addition, the ability to compromise a device at this juncture in the supply chain allows attackers to potentially defeat any network defenses at the production site. Note that the integrator has requested to remain anonymous due to certain sensitivities.

#### *3.4.2 Deployment Length.*

Shodan functions as a search engine designed to identify Internet-facing devices by continuously scanning the entire Internet. The deployment period for this research is

based on two factors: previous research and an approximation of the time required to scan all IPv4 addresses. Previous ICS honeypot research deployed honeypots for 26 to 90 days, with the primary focus of identifying attacks against Internet-facing ICS devices [51, 53]. To approximate the amount of time required to scan all public IPv4 address, this research utilized the Online Internet Scanning Calculator provided by networcon.com, which bases calculations on a single TCP SYN scan for a single port, while accounting for packet size, throughput, and total target IPv4 addresses (Table 3.1).

Table 3.1: Honeypot deployment length [37].

| | |
|---|---|
| Packet Size (TCP SYN) | 20 bytes (IPv4 header) + 20 bytes (TCP header) + 14 bytes (Ethernet header) = 74 bytes |
| Throughput | 1600 packets per second |
| Packets per Probe | 1 (Single TCP SYN) probe |
| Target IPv4 Addresses | 3,706,584,832 |
| Ports | 1 port probed |

The Online Internet Scanning Calculator estimated a scan of all public IPv4 address would require 26 days and 19 hours (Equation 3.1). Previous research set the minimum and maximum values for the deployment, while the scan approximation offered an empirical calculation ensuring the devices would be scanned once at a minimum. For sufficiency, this research uses a 55 day deployment period to account for double the estimated time required to scan all IPv4 addresses and accounting for the mean honeypot deployment of previous research.

Time approximation to scan all public IPv4 addresses:

$$\frac{3,706,584,832\ IPv4\ addresses \times 1\ probe}{1,600\ \frac{packets}{second} \times 86,400\ \frac{seconds}{day}} \approx 26\ days\ 19\ hours \tag{3.1}$$

### *3.4.3 Honeypot Type.*

This research utilizes high-interaction research honeypots. As with all honeypots, the goal is to present the most realistic honeypot possible relative to the target environment and data desired. To-date, the low-interaction honeypot options do not provide an adequate level of device interaction necessary to represent the desired dataset and environment. A high-interaction honeypot provides the ability to interact with a fully functioning physical PLC device.

### *3.4.4 Design Configuration.*

A primary tenet of honeypot design is that the honeypot should strive to look like a production asset [20]. This research utilizes the Allen-Bradley ControlLogix PLC. The design is based on a generic characterization of Allen-Bradley devices identifiable via Shodan. This research uses the Allen-Bradley PLC due to its notoriety as one of Northern America's primary PLC suppliers. As of 2013, in North America Rockwell/Allen-Bradley maintained 60% to 70% of the market share in both original equipment manufacturer and end-user markets [2]. Figure 3.1 details the November 2013 market analysis by the ARC Advisory Group, a global market research firm for automation, asset management and control.

As the intent of each honeypot is to represent those devices currently identifiable via Shodan, an investigation into the Allen-Bradley devices currently indexed by Shodan is required. An inspection of the Allen-Bradley ControlLogix 1756 eWeb PLC web management service banner uncovered a basic Shodan query signature identifying 490 devices currently indexed by Shodan. A random sampling of 49 devices, roughly 10%, is used to design the honeypot configuration for this research. Each device is inspected to identify nine characteristics: CPU type, CPU firmware version, Ethernet module type, Ethernet module firmware version, naming conventions, number of modules, type of modules, chassis size, and available services. These characteristics are identified via basic

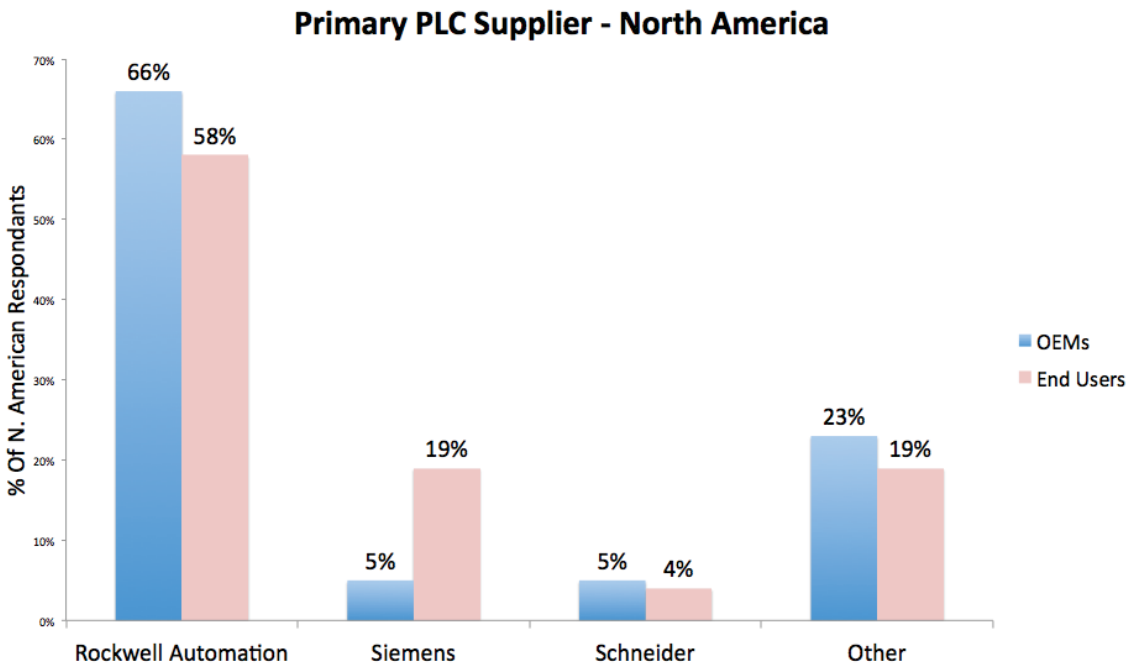


Figure 3.1: ARC market analysis of North American PLC suppliers [2].

reconnaissance of the web management console and a TCP network mapper (NMAP) scan of the device to determine the available services. A TCP scan attempts to open a connection to any available services on the target machine. The NMAP option "-p 1-65535" sets the scan to query all ports. At no point did this research connect to any secured areas of public devices, nor did reconnaissance include direct interaction with EtherNet/IP ports beyond the aforementioned service scan. Figure 3.2 depicts the web management console of a randomly selected Shodan identified device. The web management console provides the majority of device characterization data utilized to develop the honeypot design to include: device name, description, Ethernet Address, product revision, firmware version date, serial number status, and uptime. Note due to sensitivity concerns, identifiable information has been redacted in the figure.

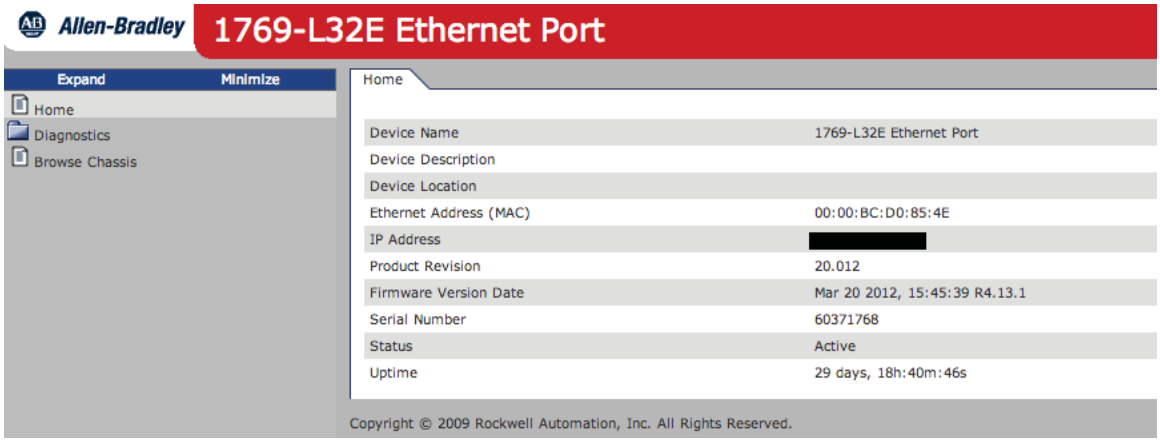


Figure 3.2: Allen-Bradley web management console - Random Shodan Sample.

Figure 3.3 depicts the browse chassis section of the web management console, identifying the associated modules and size of the chassis. Figure 3.4 shows the results of an NMAP scan of the same device and the available service: web management (port 80) and EtherNet/IP (port 44818).

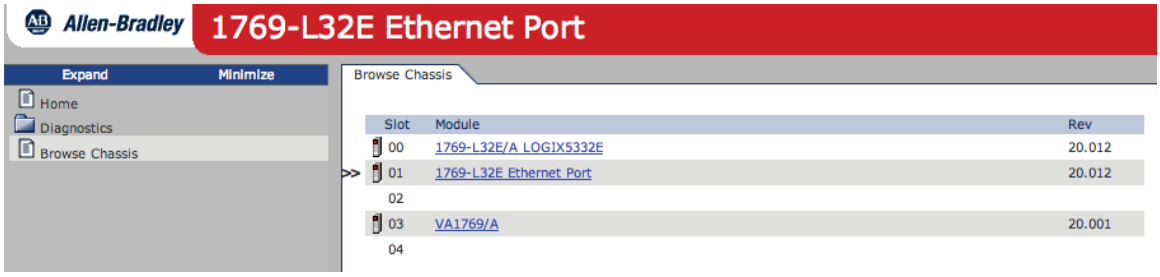


Figure 3.3: Allen-Bradley web management console - Browse Chassis.

The honeypot configuration is designed to offer the a realistic representation of Allen-Bradley ControlLogix PLCs currently identifiable via Shodan. Table 3.2 details results for each characteristic relative to the sampled dataset.

Table 3.2: Shodan Allen-Bradley PLC characteristics.

| | |
|---|---|
| CPU type | • 1756-L55 LOGIX5555<br>• 1756-L61 LOGIX5561<br>• 1756-L62S LOGIX5562<br>• 1769-L23E LOGIX5323E<br>• 11769-L35E LOGIX5335E |
| CPU Firmware version | • 13.x, 15.x,16.x, 17.x, 18.x, 19.x, 20.x |
| Ethernet Module | • 1769 Ethernet Port<br>• ENBT Ethernet Module<br>• eWeb Web Server Module |
| Ethernet Module Firmware version | • 5.xxx, 6.xxx, 9.xxx, 13.xxx |
| Naming Convention | • Control System name (e.g., 1769-L35E/A LOGIX5335E)<br>• Ethernet Module number (e.g., 1756-ENBT)<br>• Descriptor (e.g., xxx_Processing_EWEB) |
| Number of Modules | • 2, 3, 4, 5, 6, 9 |
| Type of Modules | • CPU module<br>• Ethernet<br>• I\O Communications<br>• SERCOS interface<br>• DC Input module<br>• DC Output module<br>• Output Module (Isolated Relay) |
| Chassis Size | • 5, 7, 10, 13 |
| Available Services | • Web Server (port 80)<br>• SNMP (161)<br>• EtherNet/IP (44818) |

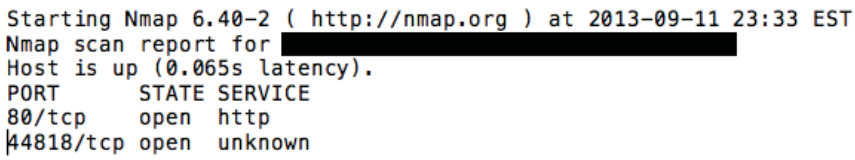


Figure 3.4: Nmap scan - random Shodan sample.

In addition to the basic characteristics cited above, it is noted that every device in the sample set is configured with a processor and Ethernet module installed in the first two available slots of the chassis. Within the sample dataset, 95% of devices are configured with the processor installed in the first slot (Slot 00) and the Ethernet module in the second slot (Slot 01).

#### *3.4.4.1 Honeypot Configuration.*

The nine characteristics provide a general device representation for the Allen-Bradley ControlLogix PLC. Based on the collected data and available resources, each honeypot is comprised of the following:

1. CPU type - Allen Bradley 1756-L61 ControlLogix Logix5561
2. Control System Firmware version - Revision 19.052.
3. Ethernet Module type - 1756-EWeb Ethernet Module.
4. Ethernet Module Firmware version- Revision 5.001.
5. Naming Convention - Descriptive (e.g., ab.2013.water.sX).
6. Number of Modules - 2.
7. Type of Modules - CPU module, Ethernet module.
8. Chassis Size - 4 Slot Chassis (CPU in slot 00 and Ethernet module in slot 01).
9. Available Services - Web Server (port 80), EtherNet/IP (44818).

Each PLC is assigned a static IP address directly accessible via the Internet. Each PLC is loaded with basic ladder logic to simulate a functioning PLC. The ladder logic

downloaded to each honeypot is a derivation of a sample temperature control application provided by Allen-Bradley RSlogix 5000, designed to take an analog input from a temperature sensor and control an analog output to a heating element. The application is altered to function on the Allen-Bradley ControlLogix 1756-L61 with firmware revision 19.052. In addition, I/O dependencies are removed and an internal application task simulates all data inputs.

Project SHINE highlighted not only the accessibility of numerous ICS devices, but also identified that numerous devices utilized weak or default authentication. As such, the web management interface is configured with default authentication settings to mimic a newly deployed PLC. Default PLC configuration allows the honeypots to assume the broadest range of targeting and represents "low hanging fruit" for exploitation. Each device is named ab.2013.water.s[honeypot number] with device description plc.water.s[honeypot number] and location site [honeypot number]. This naming convention is intended to enhance the realism of the honeypots and entice targeting by suggesting the devices are located in association with a water utility production enclave. Figure 3.5 shows one of the honeypot web management consoles detailing specifics about the device.

As shown in Figure 3.6, the honeynet design consists of four high-interaction honeypots: two standard PLCs and two PLCs with mangled banners. All PLCs are deployed Internet-facing with static IP addresses in the same subnet and configured with the same settings and default authentication. A single connection from the ICS integrator's switch is run to a 3COM Office Connect 8-port Dual Speed hub hosting the four honeypots and data collection laptop. The standard honeypot PLCs are representative of the Allen-Bradley ControlLogix devices identifiable via Shodan, while the banner mangled honeypots seek to measure the impact of the service banner data on post Shodan

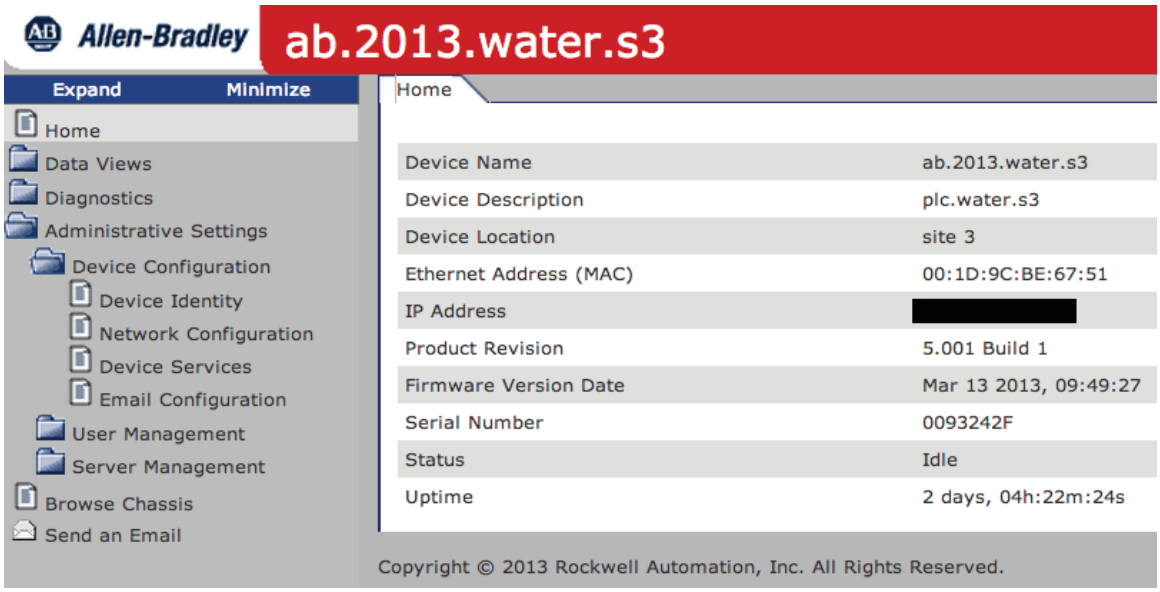


Figure 3.5: High-interaction honeypot web management interface.

identification. The honeypots also evaluate the ability to limit device exposure via banner manipulation. All honeypots expose two services to attackers:

1. HTTP (port 80) web administrative interface. The web interface is available to anyone accessing the service and requires the default credentials (administrator:null).
2. EtherNet/IP and Common Industrial Protocol (port 44818) communications protocols originally developed by Rockwell Automation for use in process control and industrial automation applications.

These represent the two most basic services offered by the Allen-Bradley ControlLogix PLC and the primary services required for device deployment within a production environment. In addition, these services represent the only two services found on every ControlLogix device from the sample set of Allen-Bradley devices currently indexed via Shodan. The web server provides access to information from the control

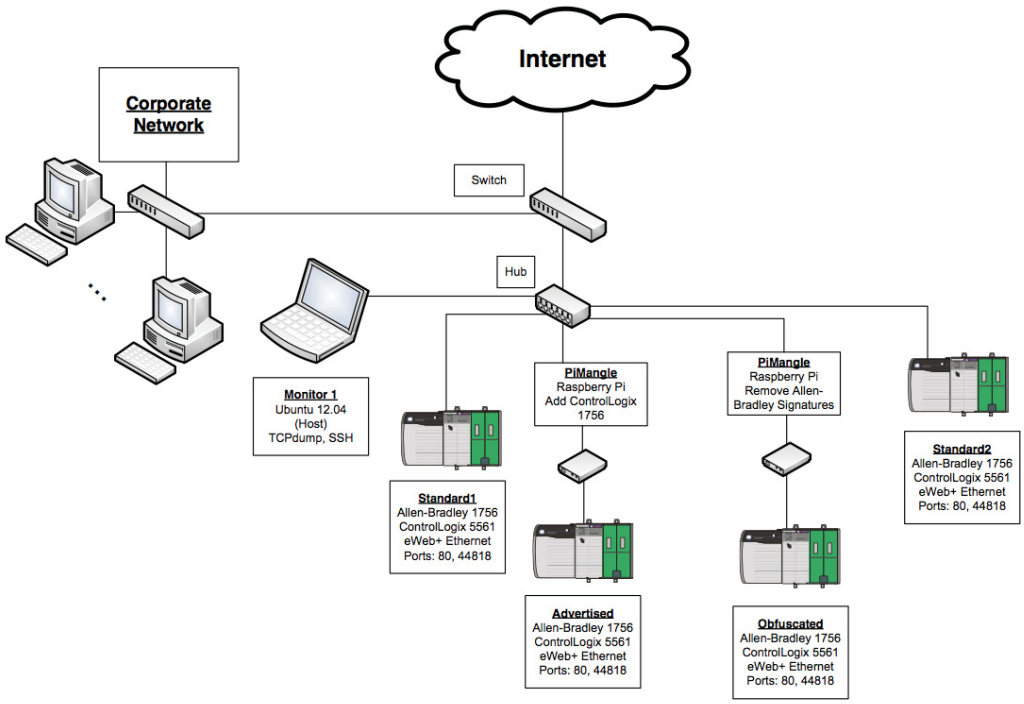


Figure 3.6: Deployment and setup co-located with ICS integrator.

system using a web browser, while also allowing remote control system monitoring and modification. The EtherNet/IP service provides remote control and management of the device. Figure 3.7 depicts the available services as noted via the web management console. Note that the two services represented as CIP are associated with one service running on port 44818. Access to this section of the web management console requires the default username and password.

Figure 3.8 depicts a NMAP service scan of the honeypots detailing open ports and services as 80 and 44818, while also detailing with 93% accuracy the device as a Rockwell Automation 1769-L23E-QB1 PLC based on NMAP's device fingerprinting service. Note the scan for all four honeypots yielded the same results. NMAP device fingerprinting is based on a comparison of device responses to specific TCP/IP probes and open TCP or UDP ports. An investigation into the NMAP fingerprint database reveals that the only

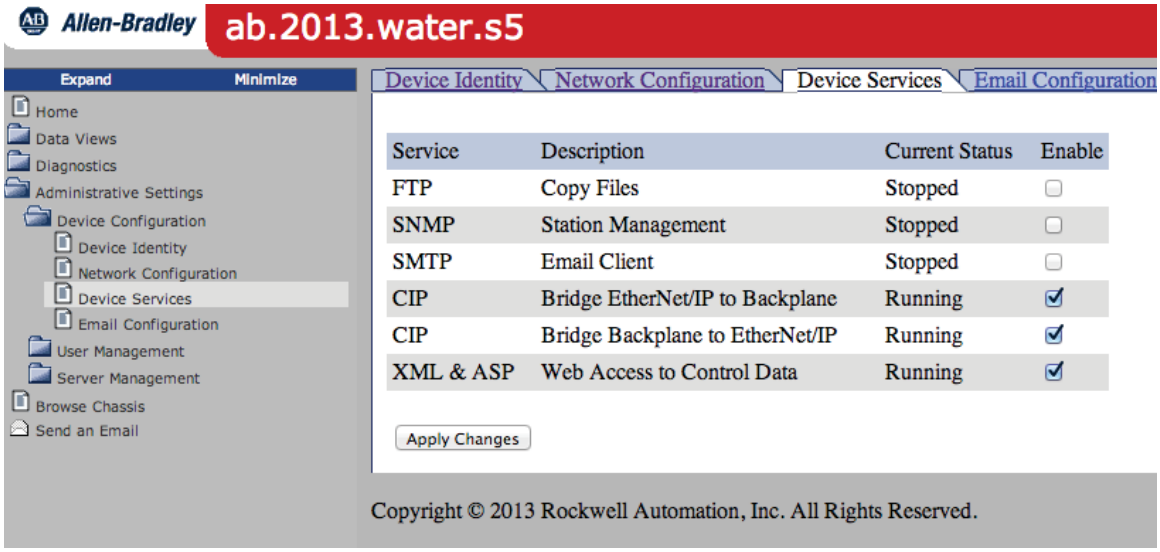


Figure 3.7: Available services provided by honeypots viewed via web management console.

Allen-Bradley device is the Allen-Bradley 1769-L23E, explaining misidentification of the Allen-Bradley model. While the NMAP fingerprint miscategorizes the model number, it presents the attacker with an indication of the device type (i.e., Allen-Bradley) and information pertinent to begin reconnaissance (i.e., open ports and banner description).

#### *3.4.4.2 Standard Honeypots.*

Standard honeypots measure the impact of Shodan on Internet-facing ICS devices. The two standard honeypots consist of an Allen Bradley 1756-L61 ControlLogix 5561 revision 19.052 with eWeb Ethernet module revision 5.001, four slot chassis, and DC power supply. Standard honeypots are configured with a static Internet-facing IP address concurrent to the ICS integrator's corporate network. Figure 3.9 outlines the basic network setup for the standard honeypots, wherein the standard honeypots consist of an Allen-Bradley ControlLogix PLC connected to the ICS integrator's switch via a hub. Note all honeypots are connected to a single hub as represented in Figure 3.6. This

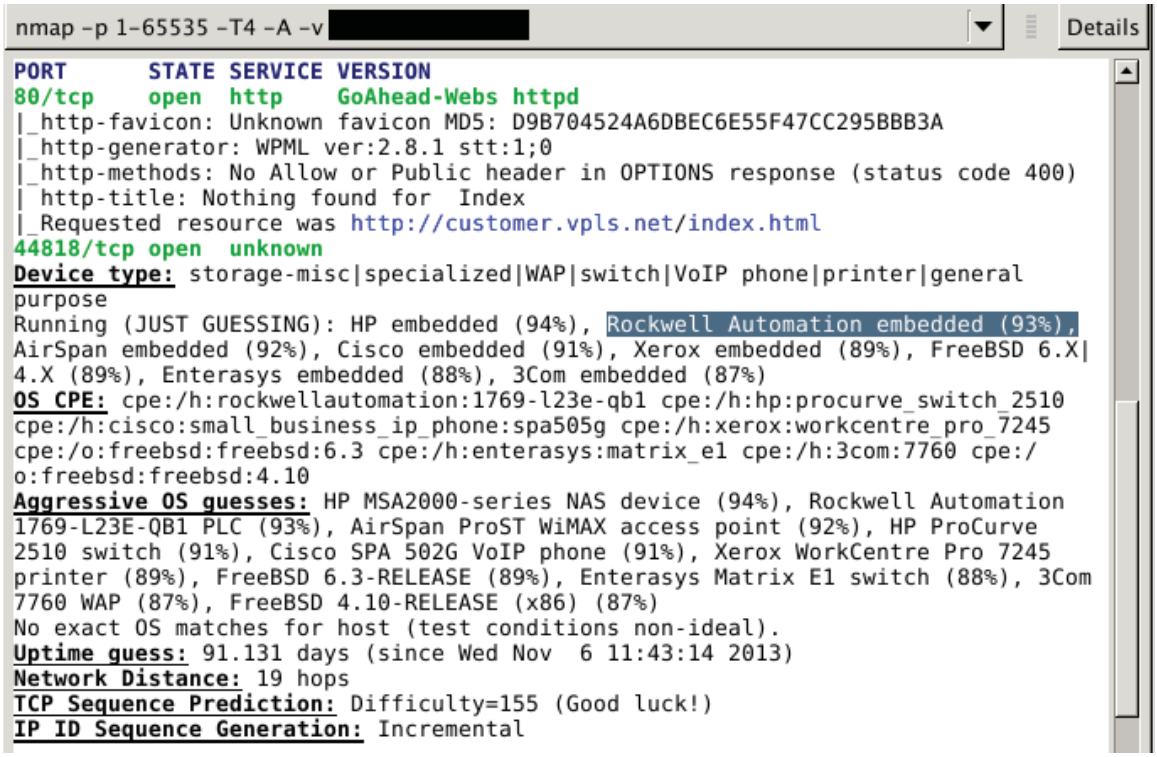


Figure 3.8: NMAP scan of Allen-Bradley ControlLogix 5561 PLC.

configuration facilitates the ability to have a single monitoring laptop conduct full packet captures for all devices.

Figure 3.10 shows the web management console homepage detailing device name, description, Ethernet Address, IP address, product revision, firmware version date, serial number status, and uptime. Standard honeypots are programmed with a basic ladder logic to simulate activity. Name resolution is enabled on each PLC and configured with public DNS servers, primary 209.244.0.3 and secondary 209.244.0.4.

#### *3.4.4.3 Banner Mangled Honeypots.*

Shodan offers users a searchable database of Internet accessible devices and any banners associated to those devices [43]. This mechanism supposes the greatest threat to device identification lies in the information revealed by device service banners. In order to

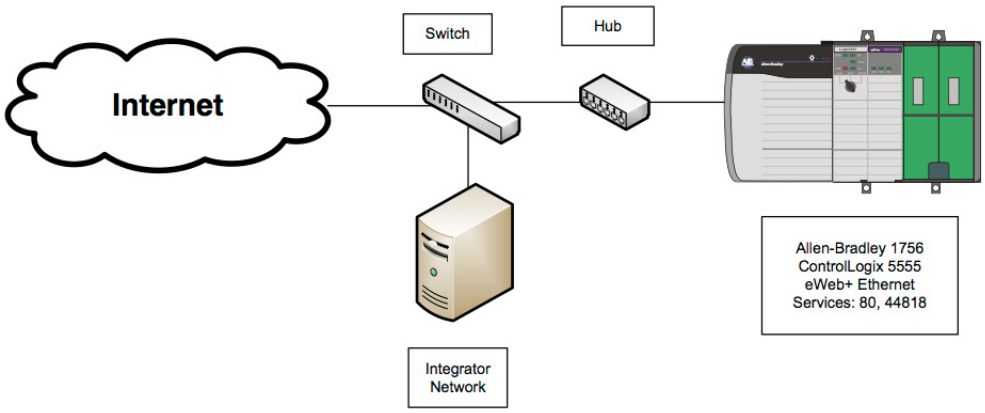


Figure 3.9: Standard honeypot design.

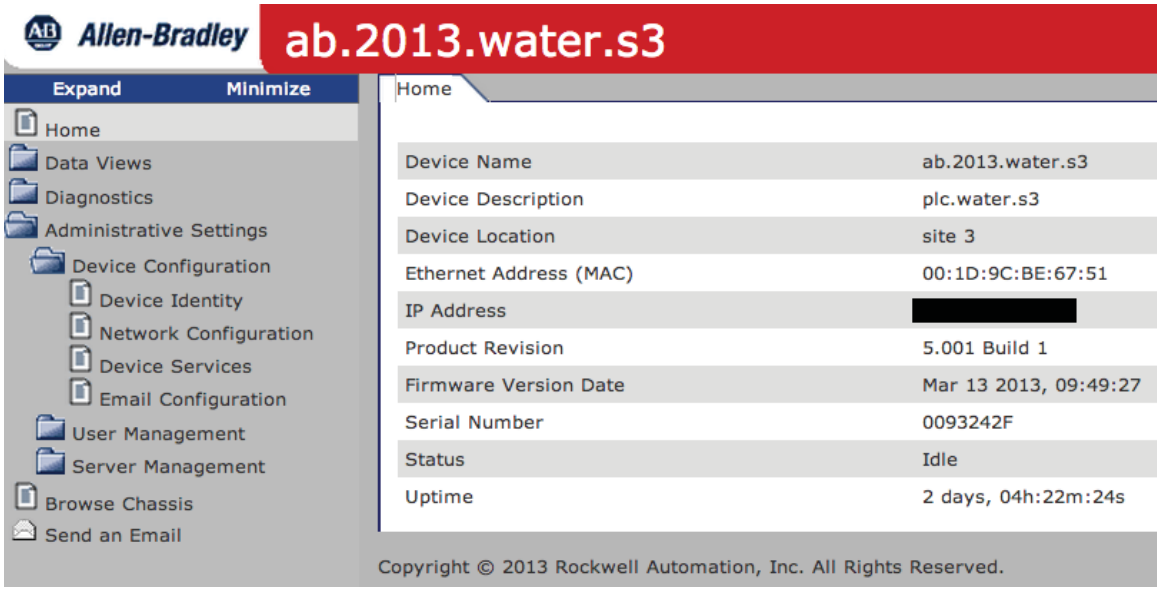


Figure 3.10: Standard honeypot web management console.

assess the impact of ICS device service banner data relative to identification within Shodan, two honeypots are deployed with altered service banners, one indicating the device model and one obfuscating the device by removing Allen-Bradley ControlLogix

PLC indicators. This design also evaluates the ability to disguise a PLC from Shodan query discovery and signature development. Both banner mangled honeypots are configured with the same specifications as the standard honeypots, consisting of an Allen Bradley 1756-L61 ControlLogix 5561 CPU with firmware revision 19.052, eWeb Ethernet module revision 5.001, four slot chassis, and DC power supply. Banner mangled honeypots are configured with static Internet-facing IP address concurrent to the standard honeypots. To manipulate the device service banner a transparent bridge is inserted between the physical PLC and the Internet, altering any outgoing service banners (Figure 3.11).

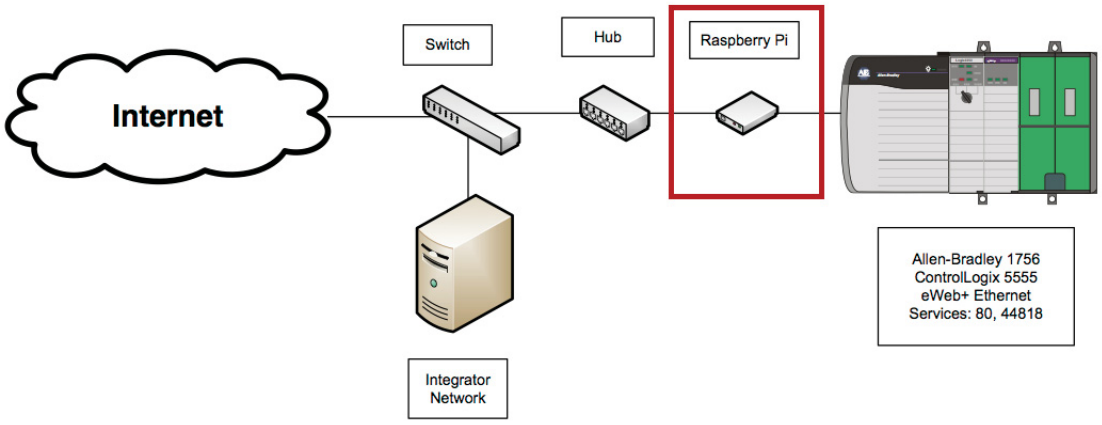


Figure 3.11: Banner mangled honeypot transparent bridge implementation.

The transparent bridge is designed using a Raspberry Pi configured with Linux IPtables and bridge-utils to bridge the on board Ethernet card and a second USB Ethernet adapter. A Raspberry Pi is a credit-card-sized single-board computer running a version of Debian Linux. There are two different banner mangled honeypot implementations: Obfuscated and Advertised. Both use Python implementations to inspect outbound traffic from the PLC and alter any packets containing specific banner data. In the case of the

Obfuscated honeypot, the default Allen-Bradley 1756-L61 ControlLogix 5561 PLC port 80 banner `Server: Go Ahead-Webs` is replaced with a random string to obfuscate the device and limit exposure. The Advertised honeypot utilizes banner mangling to make the device more readily identifiable by replacing the default banner `Server: Go Ahead-Webs` with `Allen-Bradley ControlLogix 1756`. As a result, once the PLC is indexed by Shodan, any query for Allen-Bradley, ControlLogix and/or 1756 should reveal the Advertised PLC. Figure 3.12 illustrates the comparison between the default, Obfuscated, and Advertised honeypot banner grabs.

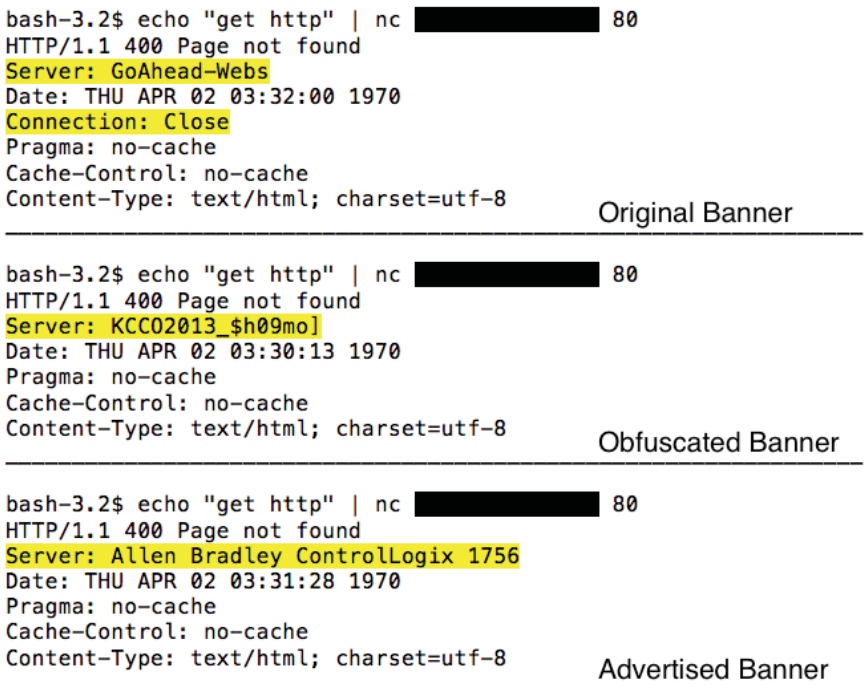


Figure 3.12: Transparent bridge banner manipulation.

### *3.4.5 Setup Validation.*

Prior to deployment, each honeypot device is evaluated in a lab environment to ensure the devices are correctly configured. The testing environment consists of a closed

private network containing a hub, the four honeypots, and a data collection laptop. Each honeypot is assessed according to NMAP scans, web server interaction, EtherNet/IP accessibility and functionality. Once online, each device is scanned using NMAP to determine available services while also obtaining a device fingerprint. Each device is then assessed via web management interface traversal to ensure all setting and configurations are accurate relative to the predetermined device design. This includes: available services, default authentication, and naming conventions. Finally, each device is tested to confirm EtherNet/IP protocols are available and the ladder logic is accessible. EtherNet/IP validation is accomplished using the Allen-Bradley RSLogix 5000 software to access the ladder logic loaded to each device, confirming the ability to upload, download, and make changes.

Once deployed, each device is again scanned via NMAP to generate a device fingerprint for comparison with the secured lab environment fingerprint. Each device is scanned daily using a Python script to conduct a banner grab of port 80, specifically recording the banner date and time. Results are stored according to IP address, date, banner date and time. If the device is unresponsive the script returns the IP address, date and a message stating "IP address X.X.X.X unresponsive." Banner grabs are conducted using the netcat command in combination with grep to filter for the service banner date and time.

#### *3.4.6 Data Collection.*

A network monitor, Dell Precision M4500 running Ubuntu 12.04, is deployed alongside the honeynet to conduct full packet captures of all traffic destined for target devices. The monitor utilizes TCPdump for packet capture and SSH for remote packet collection. TCPdump is a Linux command line packet analyzer capable of intercepting TCP/IP packets transmitted or received over the network. The network monitor laptop is connected via CAT5 cable to the network hub collecting all traffic. Ethernet hubs connect

multiple devices together acting as a single network segment. Subsequently, the network monitor collects traffic for all four honeypots. The monitoring laptop is configured with no external Ethernet IP address to prohibit identification by actors interacting with the honeypots. The monitoring laptop is configured for remote capture exfiltration via the ICS integrator wireless network. A packet capture bash script executes the TCPdump command every night at midnight, creating a capture file for each day. Below is a generic representation of the TCPdump commands used to capture network traffic for this research.

- tcpdump -i eth1 not host 1.2.3.4 -w capture.pcap

TCPdump monitors all traffic on the system interface eth1 and outputs the resulting network capture to file "capture.pcap." TCPdump was tested to determine packet accuracy and packet loss. At the end of each capture, TCPdump provides the total packets captured, packets received by filter, and packets dropped by kernel. To test TCPdump, a 1.5GB sample pcap file was obtained from Netresec.com publicly available files. TCPreplay was used to replay the pcap and TCPdump was used to capture the traffic. In this lab experiment, TCPdump indicated zero packets dropped by kernel. A study completed by the University of Michigan tested the effects of systematic packet loss on aggregate TCP flows, where in TCPdump was used to collect every packet transmitted and received. Over a two day period, the study observed 9,263 losses out of 62,379,519 total packets, yielding a loss rate of 0.01% [21]. These two experiments provide the validation for the TCPdump tool.

The backend analysis utilizes the Tshark processor provided by Wireshark to evaluate traffic destined for each individual honeypot. The command below represents a Tshark command to read in the daily packet capture and output all traffic relative to a single honeypot.

- tshark -r input.pcap -w output.pcap -R "ip.addr == 1.2.3.5"

To determine when a device is indexed via Shodan, this research utilizes the Shodan API to query Shodan for the specific host (i.e., IP address). This script queries the Shodan database twice daily. This script stores IP address, date, and the device service banner as identified by the Shodan database. If the device has not been indexed, the script returns IP address, date, and the message “no results.” Any successful query of the Shodan database is followed by a visual inspection of the Shodan web interface to determine if the device is identifiable. In addition, an inspection of the raw packet capture from the monitor laptop is conducted to correlate time and date stamp for the indexing. Figure 3.13 depicts a comparison between the raw packet identified via visual packet inspection in Wireshark and the data available via Shodan web interface.

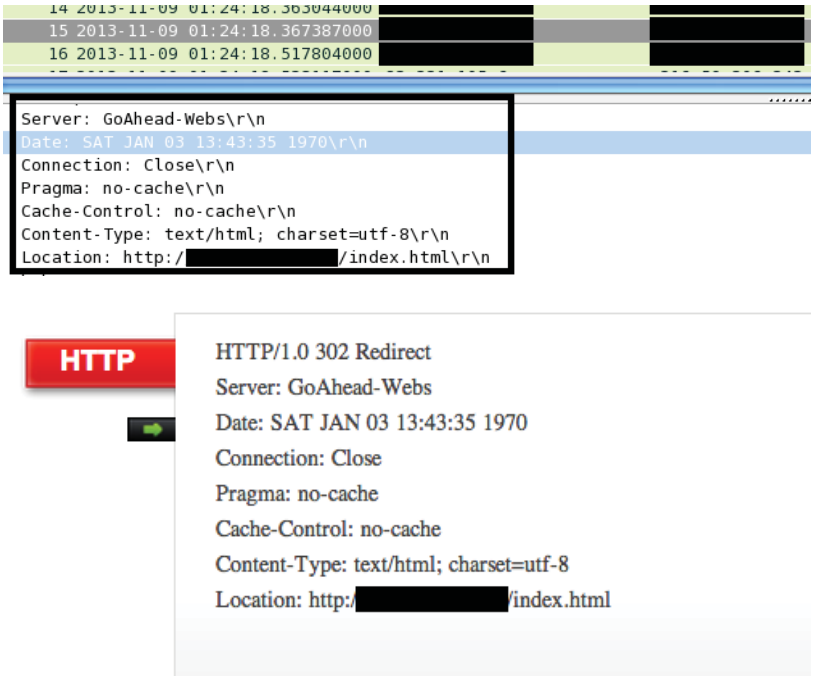


Figure 3.13: Wireshark visual packet inspection of Shodan successful device index.

## 3.5 Evaluation

This research evaluates the impact of the Shodan computer search engine on Internet-facing ICS devices by examining Shodan indexing functionality, network activity, and malicious activity targeting the honeytpots.

### *3.5.1 Indexing Functionality.*

Prior to analyzing the network activity, it is critical to define Shodan's indexing functionality. Shodan indexing functionality is evaluated by measuring the time from initial deployment to: the initial Shodan service interrogation; first successful web management index; Shodan web interface identification; and subsequent successful indexes (i.e., Shodan indexing frequency) (Table 3.3).

Table 3.3: Metrics for evaluating Shodan device scanning and indexing functionality.

| | |
|---|---|
| Indexing Functionality | 1. Initial Shodan service interrogation (days)<br>2. 1st successful web managsement banner indexing (days)<br>3. Device identifiable via Shodan web interface (days)<br>4. Shodan indexing frequency (days) |

Analysis measures the amount of time from initial deployment to the first Shodan service interrogation. In personal communications with Matherly, he revealed Shodan operates by randomly selecting an IP address, then randomly selecting a service from a set of services for interrogation. Therefore, Shodan's initial scan may interrogate a service not offered by the device resulting in an unsuccessfully index; however, this measurement marks the earliest possible opportunity for a device to be indexed.

This research then measures the amount of time a newly connected device is online before it is successfully indexed by Shodan. A successful scan is defined by the interrogation and banner grab of an available service. For this research, a successful scan

is achieved by the interrogation of the web server (port 80), as the honeypot devices are configured to offer web management and EtherNet/IP (port 44818), Shodan is not currently designed to interrogate EtherNet/IP.

This research then measures the amount of time between device deployment and the point when the device is identifiable via the Shodan web interface. Shodan offers two methods for device identification, web interface and API. When Shodan successfully scans a device, the data is compiled in the Shodan database, but is not immediately available via the web interface. Therefore, this measurement defines the point at which a device is most widely identifiable via both Shodan API and web interface, offering device identification to both sophisticated and basic users.

Finally, this research examines the frequency of Shodan indexing by recording the number of successful indexes over the 55 day deployment period. This provides insight into how frequently Internet-facing devices are scanned by Shodan.

#### *3.5.2 Network Activity.*

This research compares the network activity levels of each device as a result of Shodan identification. Network activity is defined as TCP connections, total TCP packets, and unique IP addresses interacting with the honeypot. Shodan identification is defined as the date a honeypot is first identifiable via the Shodan web interface. This delineation serves to divide network traffic for each honeypot into two datasets: pre-identification and post-identification. Each dataset (i.e., pre-identification and post-identification) is further subdivided into seven day subsets. The seven day period accounts for a standardized amount of network traffic for analysis. For the pre-identification, seven days subsets are determined by counting back from the date of web interface identification. For the post-identification, seven day subsets are determined by counting forward from the date of web interface identification. Figure 3.14 provides an example of network activity dataset determination. In this example, the honeypot was Shodan web interface identifiable on the

twentieth day of deployment. The pre-identification dataset is broken into three subsets: Pre1, Pre2, and Pre3. The post-identification dataset is broken into five subsets: Post1, Post2, Post3, Post4, and Post5.

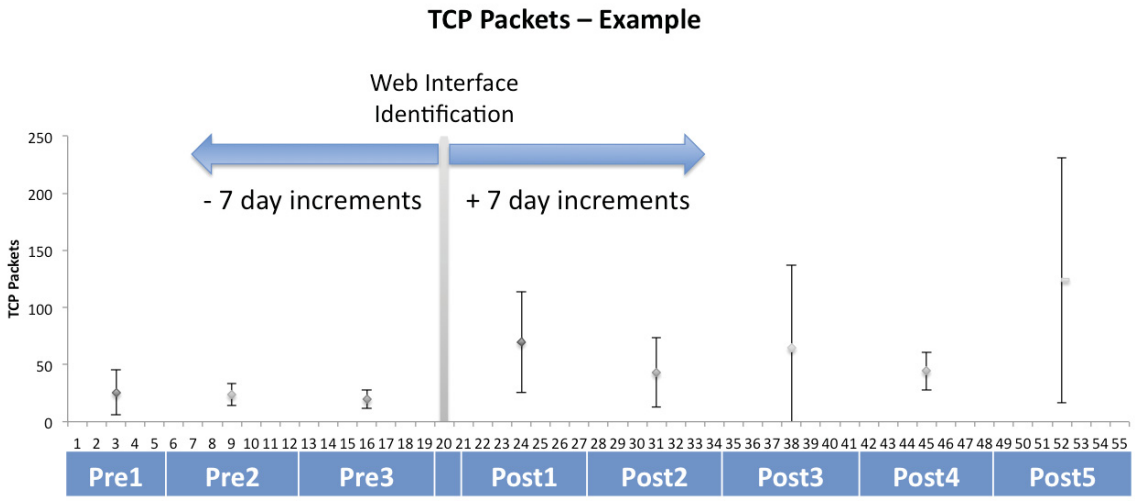


Figure 3.14: Network activity subdivision for analysis.

Network activity comparative analysis is conducted using linear trending, subset mean averages, and one-tailed pairwise t-tests. Linear trending is examined over the full 55 day deployment and offers a method of characterizing the overall change, while also quantifying the magnitude of change. Linear smoothing, also known as moving mean calculation, is used to account for variances in network activity and is calculated using a seven day moving mean. Linear trending is accompanied with an r-squared value indicating the "goodness of fit" [40]. The goodness of fit value ranges from 0 to 1, wherein an r-squared value of 1 indicates a perfect fit to the linear trend. Note the goodness of fit does not relate to the statistical significance of the trend line; statistical significance is determine by t-tests.

Analysis compares the mean averages for pre-identification subsets to the mean averages for post-identification subsets using a 95% confidence interval (e.g., Pre1 compared to Post1, Pre1 compared to Post2, Pre1 compared to Post3). A comparison of mean averages reveals if Shodan identification results in an increase in network activity. For example, if all post-identification mean averages fall above pre-identification mean averages, accounting for a margin of error represented by a 95% confidence interval, then the observed difference in network can potentially be attributed to Shodan identification.

A t-test is a statistical significance test used to determine if two sets of data are significantly different from each other [40]. This research seeks to identify any increase in activity post Shodan identification and as such utilizes a one-tailed t-test which specifically tests the relationship between two datasets in a single direction. Analysis uses a pairwise t-test, wherein each pre-identification subset is compared to every post-identification subset (e.g., Pre1 compared to Post1, Pre1 compared to Post2, Pre1 compared to Post3). Note pairwise t-tests require sample sizes of equal size, therefore pre-identification and post-identification subsets containing less than seven days will not be used for statistical significance testing. A t-test results in a p-value ranging from 0 to 1. This value indicates whether the null hypothesis should be rejected relative to the predefined confidence interval. This research hypothesizes that Shodan honeypot identification results in an increase in network activity, as such the null hypothesis is that Shodan identification "does not" result in an increase in network activity. This research uses a 95% confidence interval, therefore to reject the null hypothesis the t-test must result in a p-values less than 0.05. Table 3.4 provides an overview of the network activity metrics used to measure the impact of Shodan.

### *3.5.3 ICS Specific Targeting.*

The essence of a honeypot assumes all device interaction is malicious, as the device is non-production and should receive no legitimate traffic. It is also assumed any device

Table 3.4: Network activity evaluation metrics.

| | |
|---|---|
| Linear Trending (55 day period) | 1. TCP connections<br>2. TCP packet count<br>3. Unique IPs |
| Subset Mean Averages (Pre/Post-identification) | 1. TCP connections<br>2. TCP packet count<br>3. Unique IPs |
| Subset Pairwise T-test (Pre/Post-identification) | 1. TCP connections<br>2. TCP packets<br>3. Unique IPs |

directly connected to the Internet will receive a level of suspicious and malicious interaction and targeting. This research focuses on ICS specific device targeting and attacks as a result of Shodan indexing and identification. ICS specific targeting is defined as PLC web management server reconnaissance, unauthorized access to secured areas of the PLC web management server, any modifications or modification attempts to PLC configurations, and any interaction or specific attacks against ICS specific protocols. Analysis is accomplished by visual packet inspection and Snort IDS analysis.

***Visual Packet Inspection.*** Visual packet inspection is conducted using Wireshark, a network protocol analyzer, to reveal targeted device reconnaissance and unauthorized access to secured areas of the PLC web management server. The web server provides access to information from the control system using a web browser, while also allowing remote control system monitoring and modification. Web management server reconnaissance is defined by a manual traversal and specific site requests. The Allen-Bradley ControlLogix 1756-L61 web server provides an implementation of HTML utilizing Active Server Page (ASP) and Extensible Markup Language (XML) files. To

identify reconnaissance activity via web management interface traversal, a visual inspection of the network traffic identifies specific HTTP Get requests containing device ASP files. For example, a request for the ASP file chassisWho.asp reveals an attempt to investigate the PLC chassis, identifying the chassis size, number of modules, and specific module identification. Figure 3.15 depicts the aforementioned request in Wireshark. Table 3.5 details the specific ASP files and their relevance.

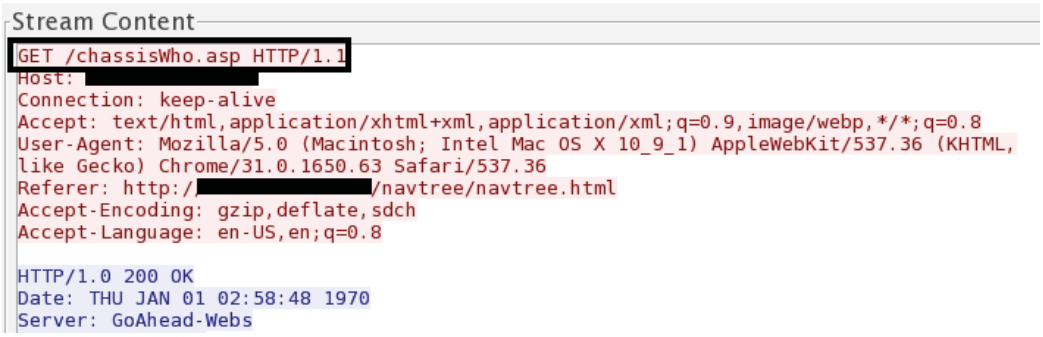


Figure 3.15: Visual packet inspection of the GET request indicating a query of the device chassis information.

The second evaluation of malicious device traversal is any attempt to access the secured areas of the web management console, both successful and unsuccessful. Each honeypot is configured with default security setting, meaning the username and password is administrator:null. A simple Google search for Allen-Bradley ControlLogix default authentication reveals the default username and password. A visual packet inspection reveals authentication attempts, as well as the username. Figure 3.16 reveals a successful login using the default credentials. This stream reveals the username “administrator” and an attempt to access the secured device identity web page.

***Snort IDS Analysis.*** To identify and enumerate specific ICS targeting, this research utilizes an implementation of the Security Onion, a Linux distribution designed for

Table 3.5: Visual Packet Inspection - device traversal.

| | |
|---|---|
| Homepage | • home.asp |
| Device Diagnostics | • diagover.asp<br>• diagnetwork.asp<br>• msgconnect.asp<br>• etherstats.asp |
| Device Configuration (Secured) | • identity.asp<br>• network.asp<br>• services.asp<br>• emailConfig.asp |
| User Management (Secured) | • editusers.asp<br>• editlimits.asp |
| Server Management (Secured) | • webManage<br>• webTime<br>• backupRestore.html<br>• serverlog.asp |
| Chassis Identification | • chassisWho.asp |

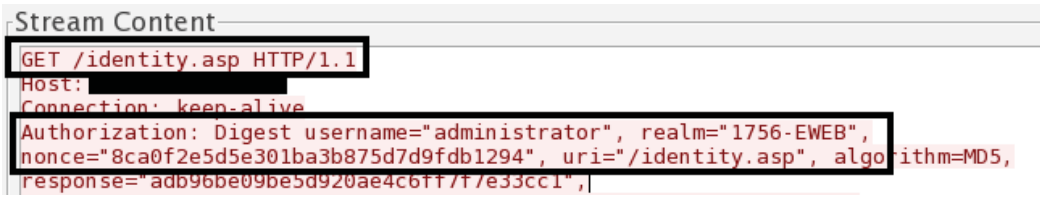


Figure 3.16: Attempt to access secured areas of the PLC web management console.

intrusion detection, network security monitoring, and log management. Tools include Snort, Snorby, Squil, netcat, and TCPreplay [4]. In addition to the latest Snort

implementations, this research utilized Digital Bonds Quickdraw SCADA IDS signatures which include DNP3, EtherNet/IP, Modbus TCP, and vulnerability signatures. Digital Bond also developed SCADA IDS preprocessors and associated Plugins for the Snort IDS which prepare the control system protocols and communication for analysis by Snort rules [13]. The SCADA preprocessors are designed to account for control system protocol fragmentation and protocol state issues, extracting message objects that can be analyzed using SCADA payload detection rule options in Snort rules [13]. Although this research focuses on the Allen-Bradley PLC and EtherNet/IP protocol, any detection of incidents against Modbus or DNP3 are of interest. This research is scoped to ICS devices, therefore any alerts or malicious activity is categorized as: (i) specifically targeting ICS devices or (ii) indiscriminate targeting of Internet-facing web servers. Snort alerts are based on the level of priority: high, medium, or low. Targeted ICS attacks are associated with high priority alerts as these are an indication of direct device scanning, automated tool exploitation, privilege escalation, unauthorized device access, and unauthorized read/write requests to a PLC [13]. Alternatively medium and low alerts are indicative of indiscriminate Internet-facing device targeting. Comparative analysis is conducted on total Snort alerts using linear trending, subset mean averages, and one-tailed pairwise t-tests as an additional measurement of Shodans impact on Internet-facing ICS device security. Table 3.5 provides an overview of the ICS specific targeting metrics.

### *3.5.4 Banner Impact.*

This research evaluates the impact of the data revealed by the device service banner relative to device identification via Shodan by measuring the level of specific ICS attacks post Shodan identification. Comparative analysis is conducted between the Advertised and Obfuscated honeypots to measure the ability to limit device exposure via banner manipulation. In addition, this research utilizes an independent party to attempt to identify Allen-Bradley ControlLogix PLCs using Shodan, with the specific intent of identifying

Table 3.6: ICS specific targeting metrics.

| | |
|---|---|
| Visual Packet Inspection | 1. Basic reconnaissance<br>2. Secured area access |
| ICS Specific Targeting (Snort IDS) | 1. ICS protocol attacks<br>2. Privilege Escalation/Unauthorized Access<br>3. Device Read/Writes |
| Comparative Analysis (Total Snort Alerts) | 1. Linear Trending (7 day smoothing)<br>2. Mean Average Alerts<br>3. Significance Testing (t-test) |

the Advertised and Obfuscated honeypots. The independent party assesses identification by two measures: no knowledge and specific Allen-Bradley ControlLogix PLC banner knowledge.

## 3.6 Summary

This research is intended to evaluate the impact of Shodan on Internet-facing ICS devices by deploying unsolicited high-interaction Internet-facing ICS honeypots. Evaluation is based on a determination of Shodan's indexing functionality, analyzing honeypot network activity levels post Shodan identification, and the enumeration of targeted ICS attacks against the honeypots. In addition, this research evaluates the ability to limit device exposure by implementing service banner manipulation.

# IV. Results and Analysis

The intent of this research is to evaluate Shodan's impact on Internet-facing ICS device security by deploying Internet-facing ICS honeypots. Each honeypot was deployed unsolicited and Internet-facing for 55 days. Data was collected from daily packet captures and analyzed using Wireshark and the Snort IDS. This chapter presents an evaluation of Shodan indexing functionality, analysis of network activity, and identification of ICS specific targeting.

## 4.1 Shodan

To understand the impact of Shodan, it is important to understand Shodan's functionality. Shodan is designed to identify any device linked to the Internet, including desktop computers, servers, printers, and web cameras.

### *4.1.1 Shodan Functionality.*

Shodan continuously scans the Internet using random functions to prevent bias of individual networks. Shodan begins by randomly generating an IP address, then randomly selecting a single service port to send a SYN scan. If the SYN scan to the random IP and random service port is successful (i.e., SYN—ACK response), Shodan initiates a banner grab and stores the resulting data in a database containing the IP address and specific banner data. If the initial SYN scan is unsuccessful, Shodan generates a new random IP and service port. Figure 4.1 presents a visual representation of the Shodan scanning routine. Note that Shodan relies on Python to conduct all device scans and port interrogation.

Shodan scans the Internet continuously and updates the database in real-time; however, data collection rates can impact the search engine. A high level of Shodan

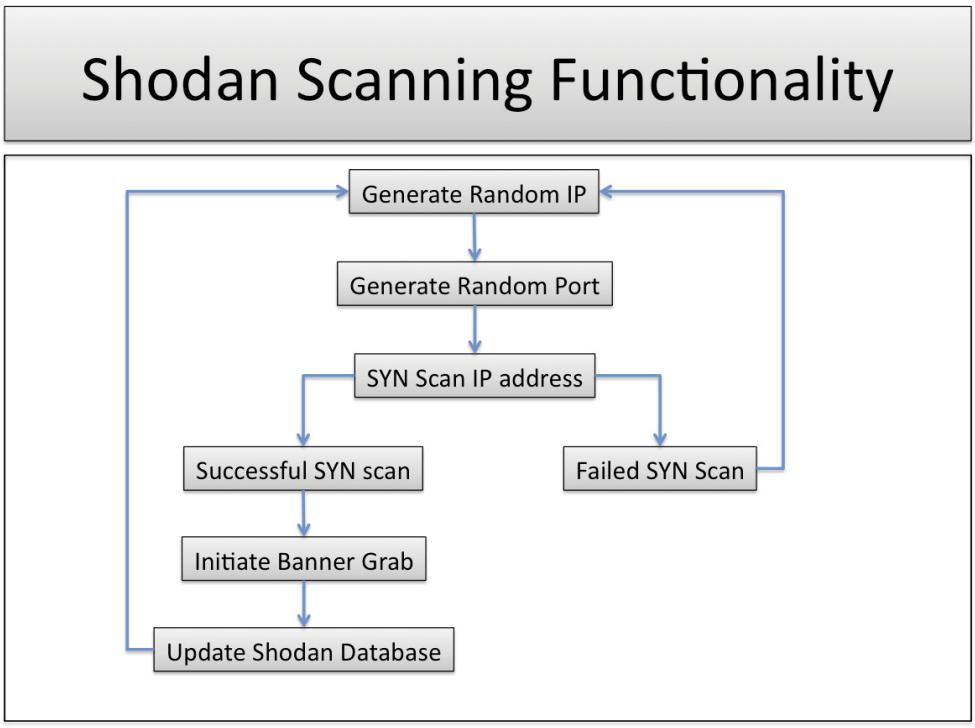


Figure 4.1: Shodan device scanning routine.

website traffic can negatively impact the search engine causing the banner update to be temporarily disabled (i.e., the ability to identify devices via web search engine queries).

Shodan offers two avenues for device identification: web interface and Shodan API. Both the web interface and API utilize the same search engine; however, the API allows utilization of the "host()" filter which bypasses the search engine and directly accesses the Shodan database. The Shodan database stores device data in "Host Profiles," providing a device summary and available services. The device summary details the IP address, Location (i.e., city and country), and latitudinal/longitudinal coordinates. Device services are listed in order of the most recent scans and port services associated with the service banner.

Shodan documentation indicates the majority of data collection occurs from standard services to include web-servers, FTP, SSH, and Telnet [43]. In total, Shodan documentation indicates 40 services interrogated during network scanning, noting the only ICS related port indexed by Shodan is SNMP (port 161). Comparing Shodan database results and raw network captures revealed a series of IP addresses conducting Shodan scans. A WHOIS lookup confirmed these addresses are associated with Shodan, indicating the owner as John Matherly. Analysis of all service interrogations initiated by Shodan IP addresses revealed seven additional undocumented services Shodan interrogates (Table 4.1). Of particular note is port 20000, which is a standard protocol port for DNP3. Communications with Shodan developers revealed the future addition of ICS specific services to include Modbus (port 502) and EtherNet/IP (port 44818).

Table 4.1: Additional Shodan service interrogation ports (i.e., undocumented).

| **Port** | **Service** |
|---|---|
| 389 | LDAP |
| 5060 | VOIP using SIP |
| 6667 | IRC |
| 9943 | ivisit Video Teleconferencing |
| 9944 | Unknown |
| 9999 | Unknown |
| 20000 | DNP3 |

### *4.1.2 Device Identification.*

Shodan attempts to index all Internet facing devices. Although not solely intended for targeting ICS devices, Shodan provides a capability to identify potential ICS devices through advanced queries. Queries are developed using a series of filters to allow users to

extract precise lists of Internet-facing devices. Results are based on information revealed by device service banner interrogation. Note that a service banner refers to information provided by a system in response to a connection request. Banner grabbing is the act of obtaining active information about a service or system through port interrogation. Figure 4.2 depicts a banner grab of an Allen-Bradley ControlLogix PLC web management console on port 80. Note that due to sensitivity concerns, identifiable information has been redacted in the figure.

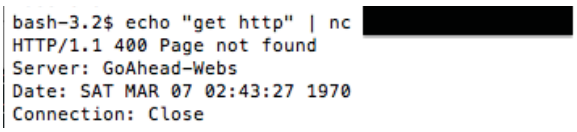


Figure 4.2: Banner grab using netcat on an Allen-Bradley PLC.

To exemplify Shodan ICS device identification, Figure 4.3 presents a generic query of Shodan for `Allen-Bradley` that identifies 98 devices. Further inspection of the details for these devices reveal they are Allen-Bradley PLC models 1747 and 1785. Both PLCs utilize SNMP with the associated SNMP service banners containing "Allen-Bradley," specifically `Allen-Bradley 1747-L553/C SLC-5/05 Series C Revision 10 1747_slc 3.46 13-Jan-06` and `Allen-Bradley 1785-L80S/C PLC5-80 Series C Revision U.2`. This illustrates Shodan search functionality as well as the impact of data revealed by service banners. Note that other Allen-Bradley PLC models, such as ControlLogix 1756, do not utilize SNMP and the service banners do not contain "Allen-Bradley," therefore they are not identified by the generic query.

This research utilizes Allen-Bradley ControlLogix PLCs. The following provides an exemplar as to specific Allen-Bradley ControlLogix PLC identification. The Allen-Bradley ControlLogix PLC web servers use a banner containing `GoAhead-Webs`

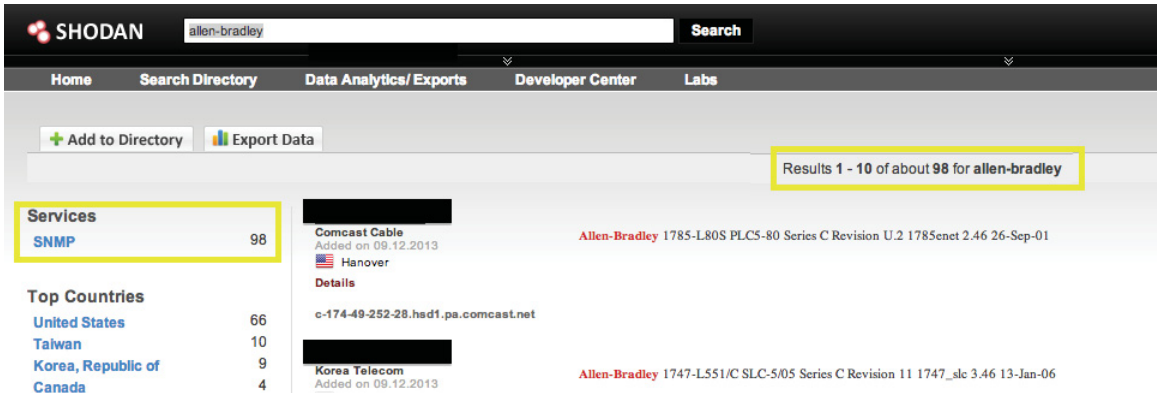


Figure 4.3: Shodan query for Allen-Bradley.

`Connection: Close index.html`. A Shodan query for that text reveals 490 Allen-Bradley ControlLogix PLC devices directly connected to the Internet (Figure 4.4). Additionally, by default the Allen-Bradley PLCs are initialized with a date of Jan 1 1970. As depicted in Figure 4.5, adding the year 1972 to the query further refines the results to 10 Allen-Bradley PLCs, which have likely been in operation for two years. As shown in Figure 4.6, an inspection of the same device revealed an uptime of 981 days, just over two and a half years.

## 4.2 Shodan Indexing

Shodan is designed to identify any Internet-facing device. The objective of evaluating Shodan's indexing functionality is to determine if, in fact, an unsolicited device will be identified and at what rate. Shodan identification can be divided into three phases: Shodan scan initialization, first successful banner grab, and web interface identification. Shodan continuously scans the Internet updating device service banners. As such, this research also examines the Shodan indexing frequency over the 55 day deployment.

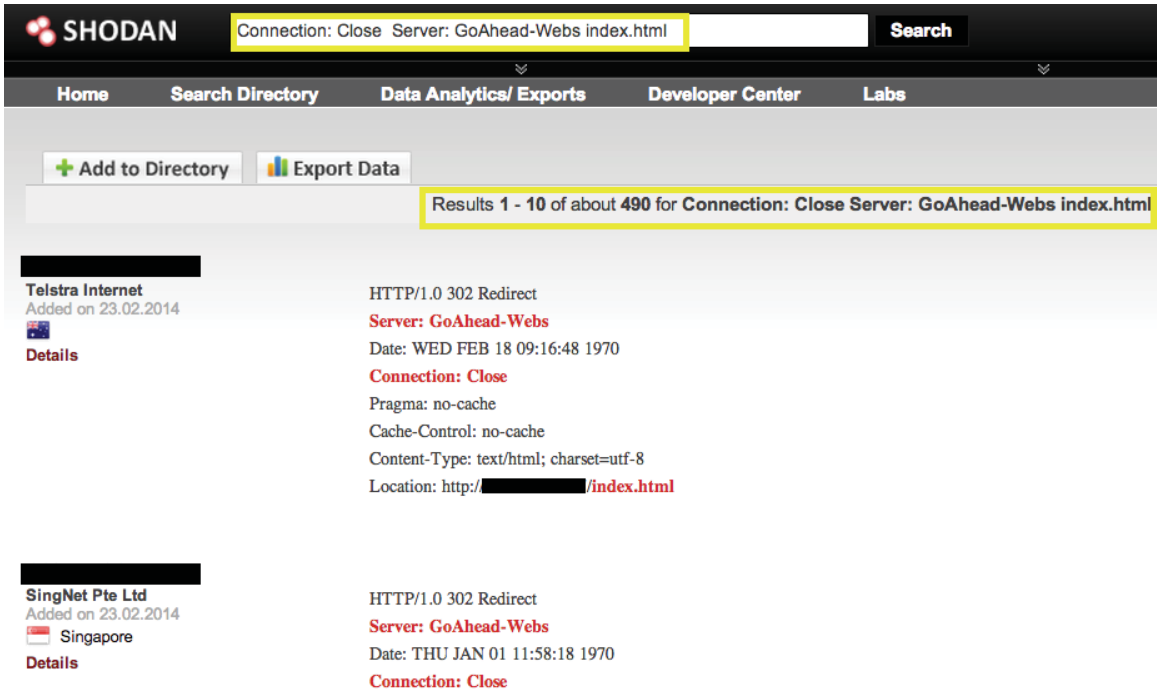


Figure 4.4: Shodan query revealing 490 Allen-Bradley ControlLogix devices.

### *4.2.1 Shodan Scan Initialization.*

Shodan operates by randomly selecting an IP address then randomly selecting a service for interrogation. As a result, the initial scan conducted by Shodan may not successfully index the device if the service is not available. Regardless, this scan marks the earliest opportunity for a device to be indexed. All four honeypots were initially scanned by Shodan in less than four days, with the Standard1 PLC and Obfuscated PLC scanned after one day. Note the initial scan against Standard2 resulted in a successful port 80 interrogation and banner grab. Table 4.2 details the number of days each device was online prior to receiving the initial service scan from Shodan, as well as detailing the initial service scanned.

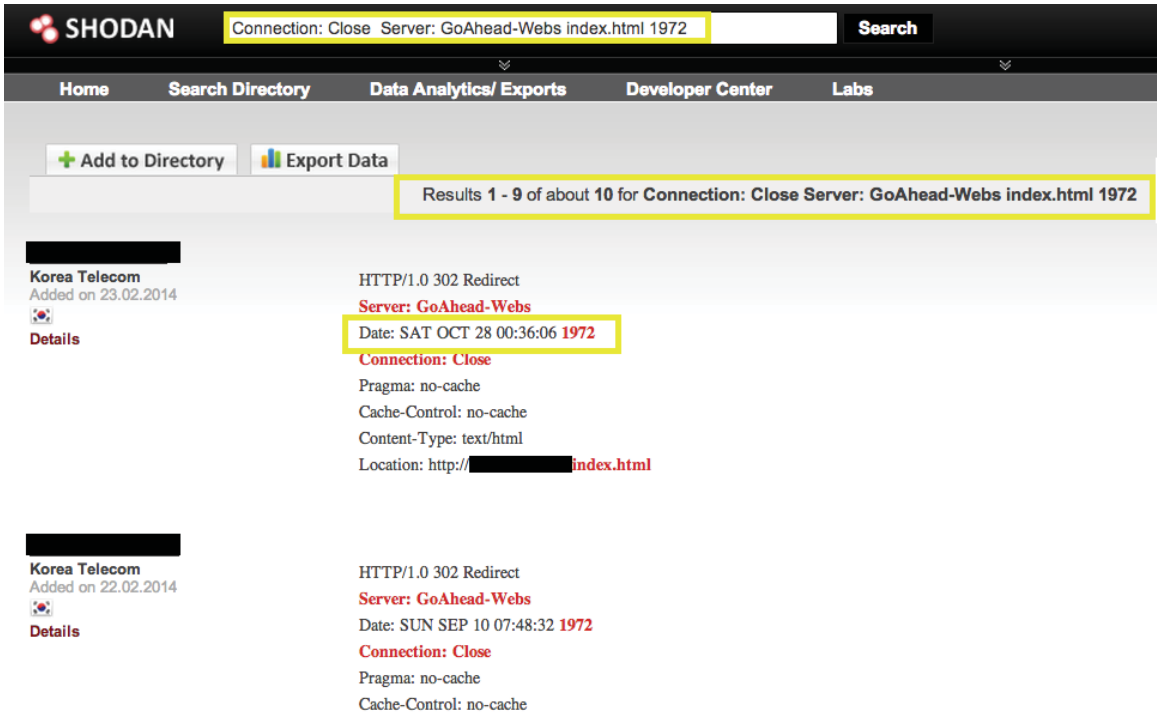


Figure 4.5: Shodan query revealing Allen-Bradley ControlLogix devices with two years in operation.

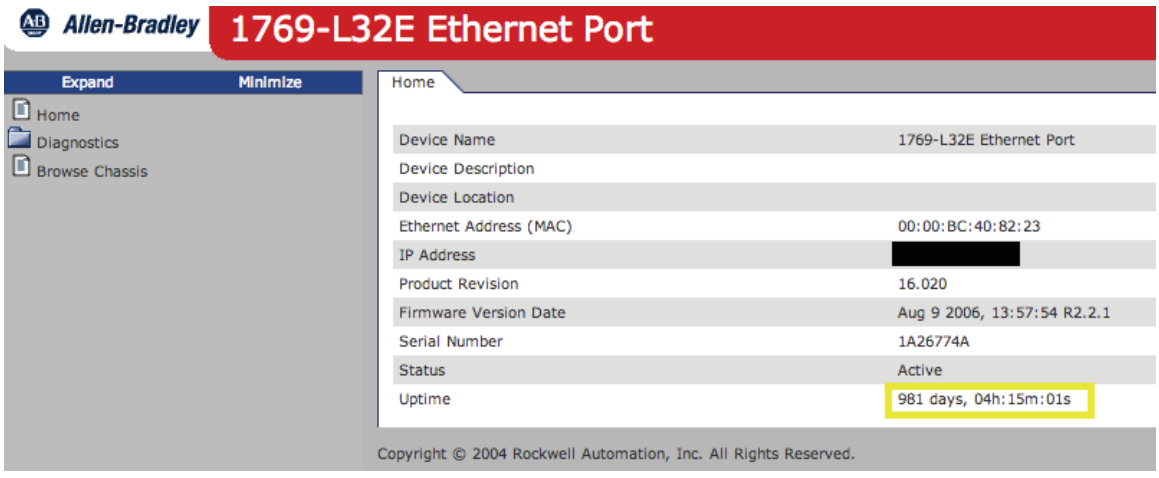


Figure 4.6: Device inspection showing an uptime of 981 days.

Table 4.2: Results for Shodan scan initialization timeline.

| Honeypot | Initial Scan | Service Scanned |
|---|---|---|
| Standard1 | 1 day (24 hrs 9 min) | 25 (SMTP) |
| Standard2 | 3 days (82 hrs 10 min) | 80 (HTTP) |
| Advertised | 3 days (85 hrs 53 min) | 25 (SMTP) |
| Obfuscated | 1 day (28 hrs 10 min) | 443 (HTTPS) |

### *4.2.2 First Successful Scan.*

Analysis examines the amount of time a newly connected device is online before it is successfully indexed by Shodan. A successful scan is defined by the interrogation and banner grab of an available service. For this research, a successful scan is achieved by the interrogation of the web server (port 80), as the honeypot devices only offer web management and EtherNet/IP (port 44818) and Shodan is not currently designed to interrogate EtherNet/IP. Shodan successfully scanned the port 80 web management service for all honeypots in less than 14 days. Table 4.3 provides details on the number of days from initial deployment to device service interrogation and banner grab.

Table 4.3: Measurement of successful Shodan port 80 interrogation.

| Honeypot | Days Online |
|---|---|
| Standard1 | 8 days 15 hours |
| Standard2 | 3 days 10 hours |
| Advertised | 6 days 22 hours |
| Obfuscated | 13 days 10 hours |

### *4.2.3 Web Interface Identification.*

The amount of time between device deployment and the point when the device is identifiable via the Shodan web interface was also evaluated. Shodan offers two methods for device identification, web interface and API. When Shodan successfully scans a device the data is compiled in the Shodan database, but is not immediately available via the web interface. Therefore, this measurement defines the point at which a device is identifiable via both Shodan API and web interface, offering device identification to both sophisticated and basic users. All four honeypots were identifiable via the Shodan web interface within 19 days. Table 4.4 details the amount of time from initial deployment to Shodan web interface identification for each honeypot, as well as the delta between first successful device index and Shodan web interface identification.

Table 4.4: Shodan web interface identification.

| **Honeypot** | **Days Online** | **Delta** |
|---|---|---|
| Standard1 | 18 days 15 hours | 10 days |
| Standard2 | 18 days 15 hours | 15 days |
| Advertised | 6 days 22 hours | 1 day |
| Obfuscated | 14 days 10 hours | 1 day |

### *4.2.4 Scanning Frequency.*

This research measures Shodan's successful scanning frequency over the 55 day deployment period. Shodan successfully indexed each honeypot a minimum of four times, with the Standard1 honeypot receiving the most successful interrogations at eight over the entire deployment period. Table 4.5 provides details the total successful Shodan service interrogation as well as the frequency in days.

Table 4.5: Successful Shodan port interrogation frequency.

| Honeypot | Total Banner Grabs |
|---|---|
| Standard1 | 8 |
| Standard2 | 6 |
| Advertised | 4 |
| Obfuscated | 4 |

#### *4.2.5 Analysis.*

No specific scanning trends were identified due likely to the random nature of Shodan indexing routine. None the less, each honeypot was successfully indexed and identifiable via the Shodan web interface within 19 days of initial deployment. In addition, each honeypot was successfully indexed via the Shodan scanning routine within 14 days, with the Standard1 honeypot indexed in eight days, the Standard2 honeypot indexed in six days, and both the Advertised and Obfuscated honeypots indexed within four days. Finally, each honeypot was initially scanned by Shodan in under four days, marking the earliest time a device could ostensibly be successfully indexed and identifiable via Shodan.

### 4.3 Network Activity

Network activity is evaluated to determine if activity levels increase post Shodan identification. An increase in network activity post Shodan identification provides indications that Shodan impacts Internet-facing ICS security. Network activity is defined as TCP connections, total TCP packets, and uniques IP addresses interacting with the honeypot. Shodan identification is defined as the date a honeypot is first identifiable via the Shodan web interface. This delineation serves to divide network traffic for each honeypot into two datasets: pre-identification and post-identification. Each dataset (i.e.,

pre-identification and post-identification) is further subdivided into seven day subsets. The seven day period accounts for a standardized amount of network traffic for analysis. For the pre-identification, seven days subsets are determined by counting back from the date of web interface identification. For the post-identification, seven day subsets are determined by counting forward from the date of web interface identification. Comparative analysis is conducted using linear trending, subset mean averages, and one-tailed pairwise t-tests.

### *4.3.1 Linear Trending.*

Linear trending is examined over the full 55 day deployment to characterize the overall change in network activity, as well as quantify the magnitude of change. Linear trending is accompanied with an r-squared value indicating the "goodness of fit." The r-squared value ranges from 0 to 1, wherein an r-squared value of 1 indicates a perfect fit to the linear trend. A positive linear trend indicates an overall increase in network activity, while a negative trend indicates a drop in overall network activity. As shown in Figure 4.7, the Standard1 honeypot saw positive linear trends across all three metrics; however, the small r-squared values as shown in Table 4.6 indicate the trend does not fit the data. As shown in Figure 4.8, for the Standard2 honeypot, the only positive linear trend occurs with respect to the number of unique IPs; however, the r-squared values across all three metrics provide an indication the trend lines are not a good fit to the actual data. The Advertised

Table 4.6: Linear trending - "Goodness of Fit" measurement (r-squared values).

| **Honeypot** | **TCP Connections** | **TCP Packets** | **Unique IPs** |
|---|---|---|---|
| Standard1 | 0.33022 | 0.58008 | 0.15437 |
| Standard2 | 0.00133 | 0.10187 | 0.03261 |
| Advertised | 0.11171 | 0.24007 | 0.00112 |
| Obfuscated | 0.00093 | 0.14169 | 0.00013 |

honeypot saw positive trending in both TCP packets and unique IPs, but a negative trend for TCP connections (Figure 4.9). The associated r-squared values for all three metrics indicated a poor fit. The Obfuscated honeypot also saw positive trending in both TCP packets and unique IPs, but a negative trend for TCP connections with associated r-squared values indicating poor fit (Figure 4.10). While in some cases linear trending indicated a positive increase in network activity over the 55 day deployment period, the associated r-squared values indicate the actual data does not fit the trend. The findings indicate that there is no linear trend associated with an increase in network activity post Shodan identification.

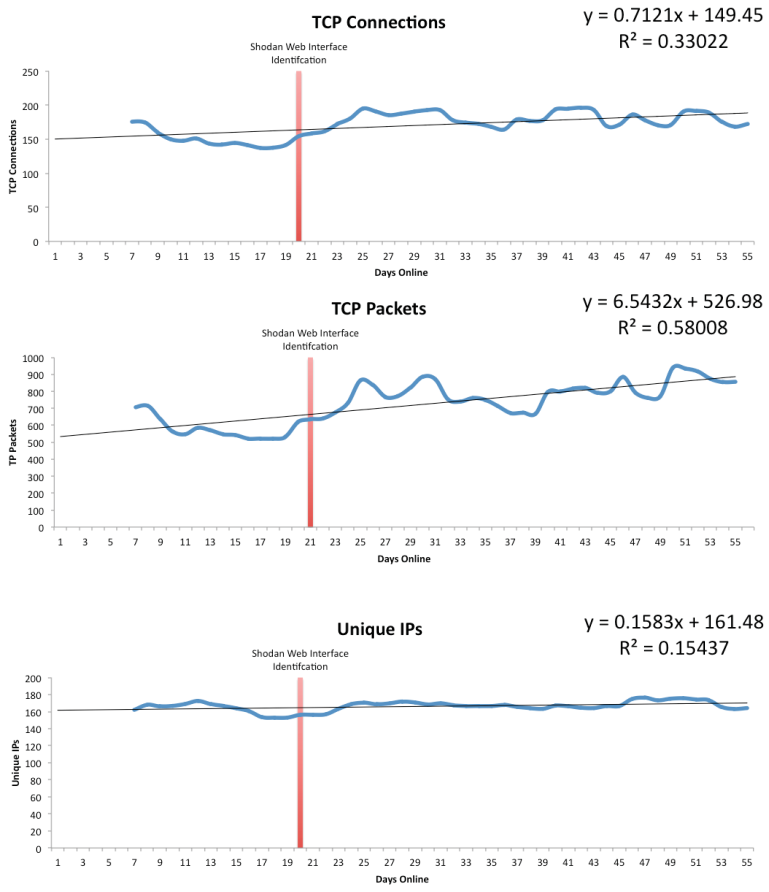


Figure 4.7: Linear trending for Standard1 honeypot - 7 day moving mean.

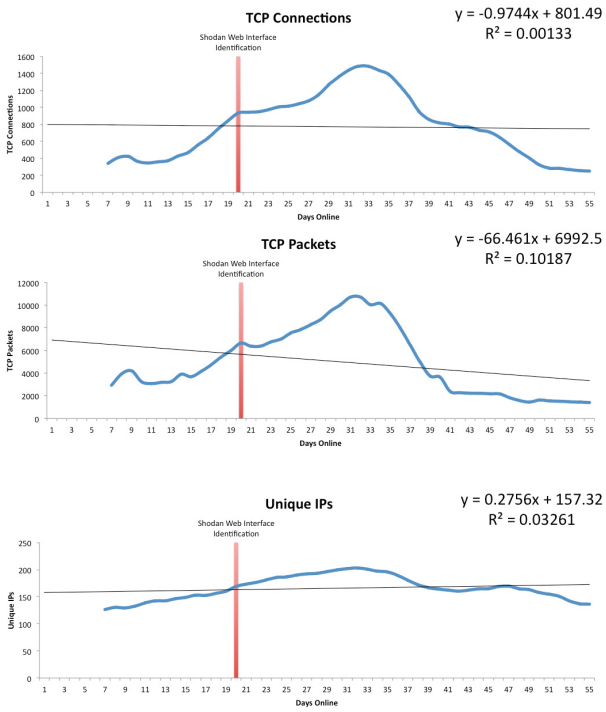


Figure 4.8: Linear trending for Standard2 honeypot - 7 day moving mean.

### *4.3.2 Subset Mean Averages.*

A comparison of the mean averages is designed to identify statistical differences in pre-identification network activity as compared to post-identification network activity levels. The mean averages for each subset are evaluated using a 95% confidence interval. Figure 4.11 details the TCP connection, Figures 4.12 details the total TCP packet, and Figure 4.13 details the unique IP mean averages for each honeypot. Save for the Standard2 honeypot subset Post2, the mean averages for post identification are not above the 95% confidence intervals for pre-identification averages, indicating the network activity did not change in comparison to pre-identification. The Standard2 subset Post2 dataset for TCP connections, TCP packets, and unique IPs are the only data points that indicate an increase in mean averages. Evaluation of this dataset revealed a series of automated scans which accounted for the increased activity.

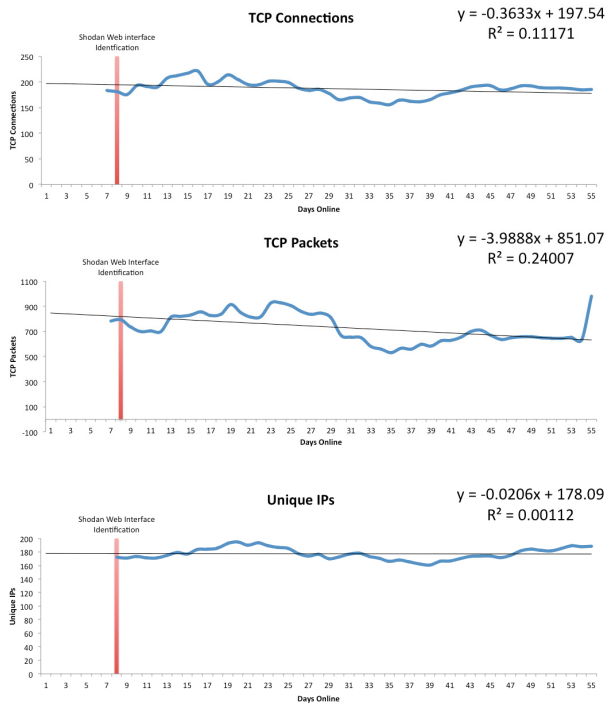


Figure 4.9: Linear trending for Advertised honeypot - 7 day moving mean.

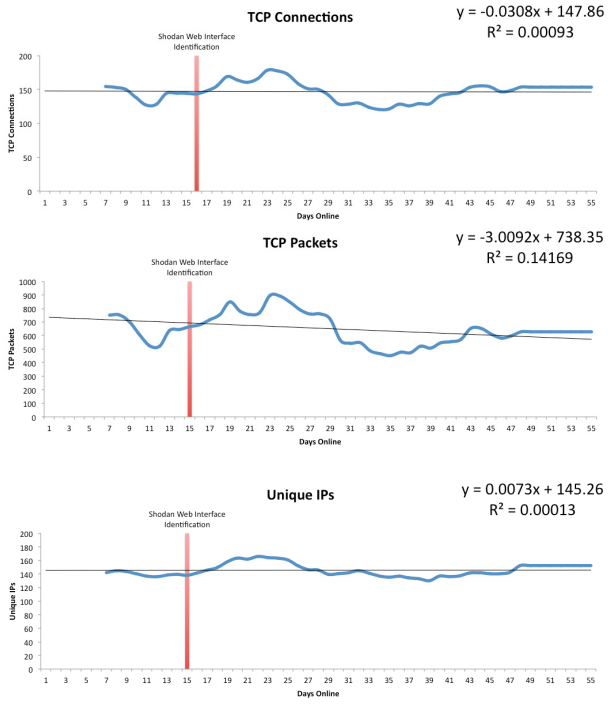


Figure 4.10: Linear trending for Obfuscated honeypot - 7 day moving mean.

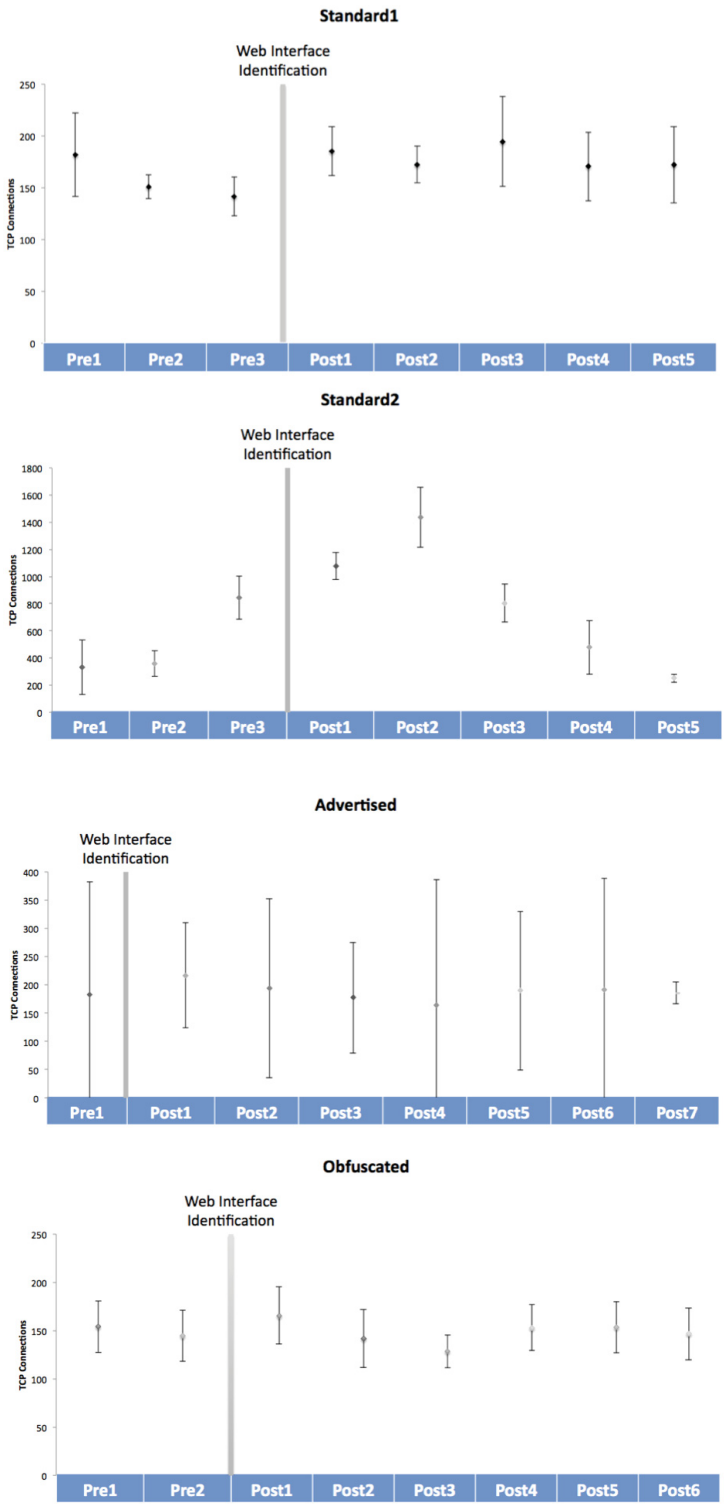


Figure 4.11: TCP connections - subset mean averages (95% Confidence Interval).

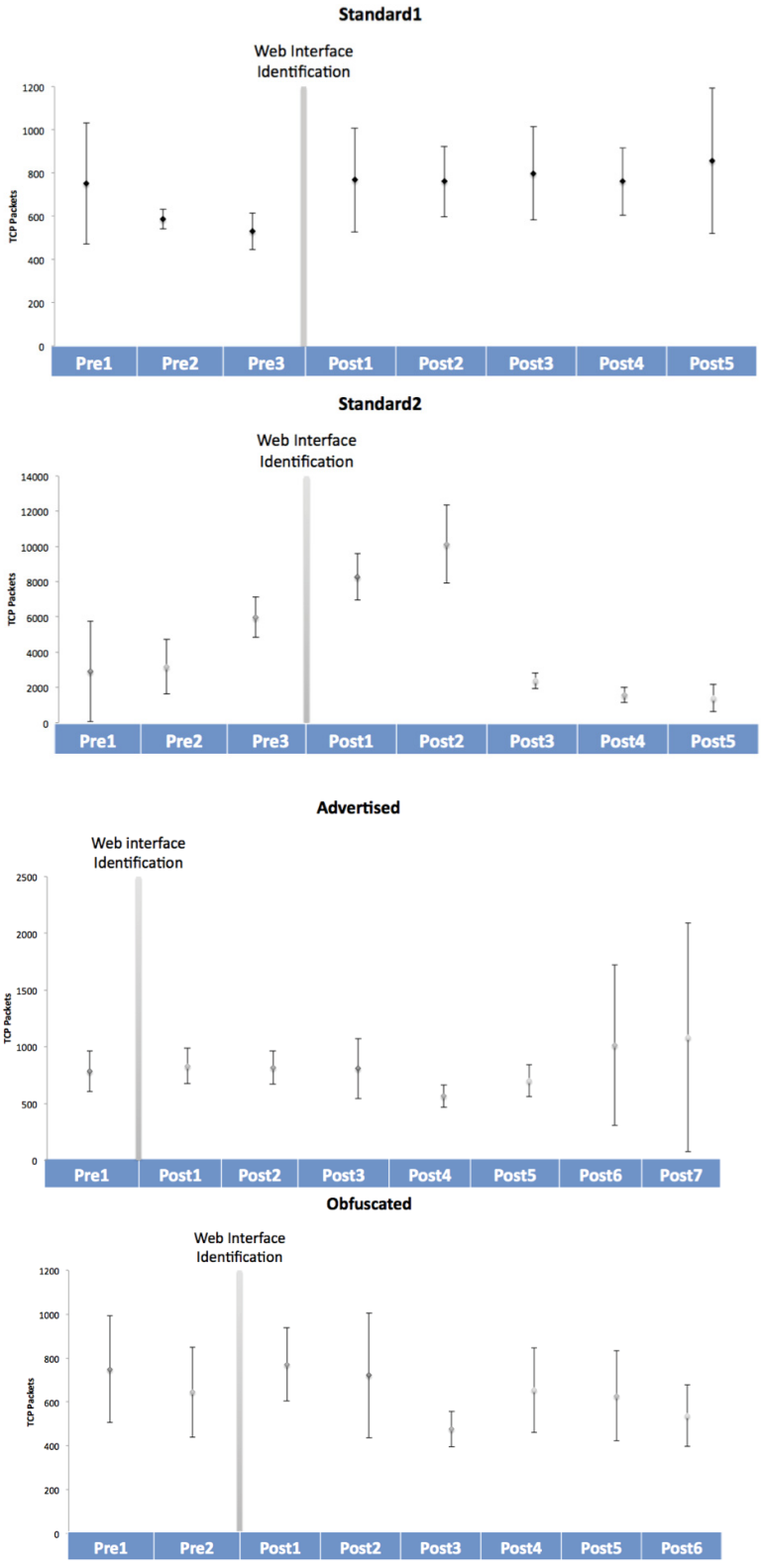


Figure 4.12: TCP packets - subset mean averages (95% Confidence Interval).

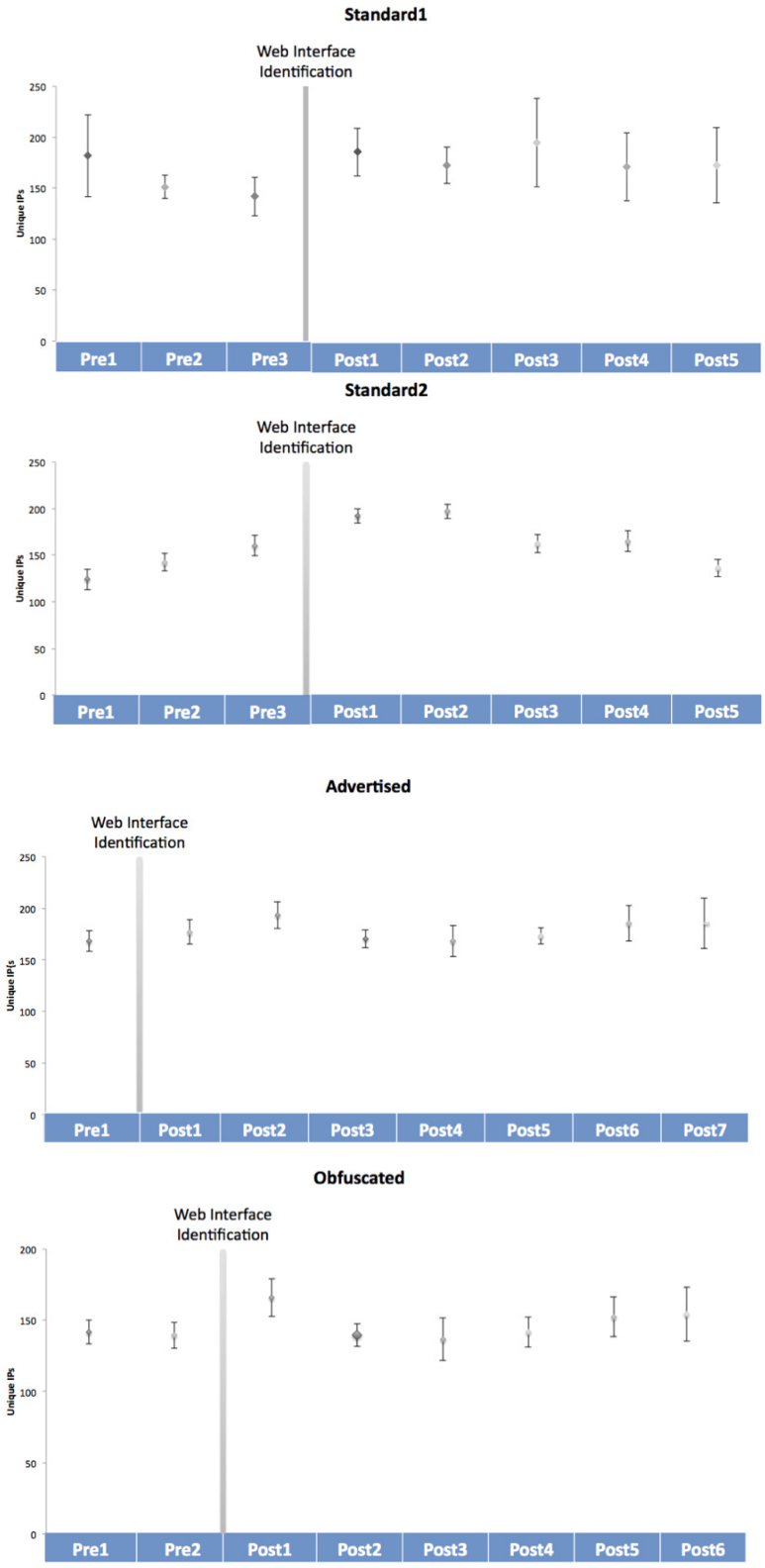


Figure 4.13: Unique IPs - subset mean averages (95% Confidence Interval).

### *4.3.3 T-test.*

A one-tailed pairwise t-test is run across each data subset to validate statistical differences in the mean average comparative analysis. Note pairwise t-tests require sample sizes of equal size, therefore pre-identification and post-identification subsets containing less than seven days are not used for statistical significance testing (e.g., Standard1 subset Pre1 contains 5 days). Table 4.7 details the Standard1 honeypot, Table 4.8 details the Standard2 honeypot, Table 4.9 details the Advertised honeypot, and Table 4.10 details the Obfuscated honeypot. A p-value of less than 0.05 indicates that the null hypothesis, Shodan does not increase network activity levels post identification, should be rejected within a 95% confidence interval. These values are indicated in bold within the tables. As the table indicates, 60% of comparisons the p-value was over 0.05 indicating no statistical difference in network activity post Shodan identification. In addition, the range of p-values is so great that no definitive indication can be discerned relating to an increase in activity.

### *4.3.4 Analysis.*

Comparative analysis did not reveal any statistical evidence supporting an increase in network activity levels post Shodan device identification. While in some cases linear trending indicated a positive increase in network activity over the 55 day deployment period, the associated r-squared values indicate the actual data does not fit the trend. A comparison of subset mean averages revealed post identification mean averages are not above the 95% confidence intervals for pre-identification averages, indicating no change in network activity levels. Pairwise t-tests were run across each dataset to validate subset mean average analysis and in 60% of the comparisons the p-value was over 0.05 indicating no statistical difference in network activity post Shodan identification. In addition, the range of p-values is so great that no definitive indication can be discerned relating to an increase in activity.

Table 4.7: Standard1 honeypot pairwise t-test results.

TCP Connections

| | Post1 | Post2 | Post3 | Post4 | Post5 |
|---|---|---|---|---|---|
| Pre2 | **0.02695** | 0.06963 | 0.07127 | 0.16176 | 0.17383 |
| Pre3 | **0.02142** | **0.03479** | **0.02133** | 0.13273 | 0.11829 |

TCP Packets

| | Post1 | Post2 | Post3 | Post4 | Post5 |
|---|---|---|---|---|---|
| Pre2 | 0.11129 | **0.03727** | 0.06228 | 0.05856 | 0.08358 |
| Pre3 | 0.06541 | **0.02765** | **0.03070** | **0.03474** | 0.07256 |

Unique IPs

| | Post1 | Post2 | Post3 | Post4 | Post5 |
|---|---|---|---|---|---|
| Pre2 | 0.28081 | 0.15482 | 0.15495 | 0.46184 | **0.04326** |
| Pre3 | **0.02172** | **0.04135** | **0.01050** | 0.06323 | **0.02225** |

Table 4.8: Standard2 honeypot pairwise t-test results.

TCP Connections

| | Post1 | Post2 | Post3 | Post4 | Post5 |
|---|---|---|---|---|---|
| Pre2 | **0.00003** | **0.00001** | **0.00102** | 0.13262 | **0.03114** |
| Pre3 | **0.00140** | **0.00390** | 0.38386 | **0.03632** | **0.00033** |

TCP Packets

| | Post1 | Post2 | Post3 | Post4 | Post5 |
|---|---|---|---|---|---|
| Pre2 | **0.00167** | **0.00024** | 0.20799 | **0.04101** | **0.04169** |
| Pre3 | **0.01145** | **0.00707** | **0.00073** | **0.00027** | **0.00008** |

Unique IPs

| | Post1 | Post2 | Post3 | Post4 | Post5 |
|---|---|---|---|---|---|
| Pre2 | **0.00003** | **0.00020** | **0.01344** | **0.00560** | 0.21718 |
| Pre3 | **0.00006** | **0.00190** | 0.41583 | 0.31875 | **0.02044** |

Table 4.9: Advertised honeypot pairwise t-test results.

TCP Connections

| | Post1 | Post2 | Post3 | Post4 | Post5 | Post6 |
|---|---|---|---|---|---|---|
| Pre1 | 0.09959 | 0.24874 | 0.38798 | 0.05854 | 0.31123 | 0.28230 |

TCP Packets

| | Post1 | Post2 | Post3 | Post4 | Post5 | Post6 |
|---|---|---|---|---|---|---|
| Pre1 | 0.34763 | 0.36775 | 0.44635 | **0.01695** | 0.21874 | 0.28703 |

Unique IPs

| | Post1 | Post2 | Post3 | Post4 | Post5 | Post6 |
|---|---|---|---|---|---|---|
| Pre1 | 0.17683 | **0.00036** | 0.41364 | 0.49536 | 0.23223 | **0.02000** |

Table 4.10: Obfuscated honeypot pairwise t-test results.

TCP Connections

| | Post1 | Post2 | Post3 | Post4 | Post5 |
|---|---|---|---|---|---|
| Pre2 | 0.13972 | 0.28200 | 0.08032 | 0.46772 | 0.48205 |
| Pre3 | 0.21266 | 0.45995 | 0.20423 | 0.33739 | 0.32586 |

TCP Packets

| | Post1 | Post2 | Post3 | Post4 | Post5 |
|---|---|---|---|---|---|
| Pre2 | 0.38344 | 0.44918 | **0.03827** | 0.27077 | 0.24842 |
| Pre3 | 0.23379 | 0.36489 | 0.11634 | 0.47665 | 0.45152 |

Unique IPs

| | Post1 | Post2 | Post3 | Post4 | Post5 |
|---|---|---|---|---|---|
| Pre2 | **0.00347** | 0.38105 | 0.31894 | 0.47689 | 0.07668 |
| Pre3 | **0.01389** | 0.50000 | 0.39741 | 0.36405 | 0.09396 |

### *4.3.5 Honeypot Interaction Country of Origin.*

Although not specific to research objectives, a broad range of network activity from multiple countries of origin was observed during the 55 day deployment period. In total, Chinese-associated IP addresses accounted for a majority of the activity against all honeypots, accounting for 32%. This is followed closely by the United States with 29%. Figure 4.14 provides a breakdown of the top ten countries, who, in sum, account for roughly 75% of all activity targeting the honeypots. Country origins were determined by a bulk IP lookup using MaxMind batch lookup service [32].

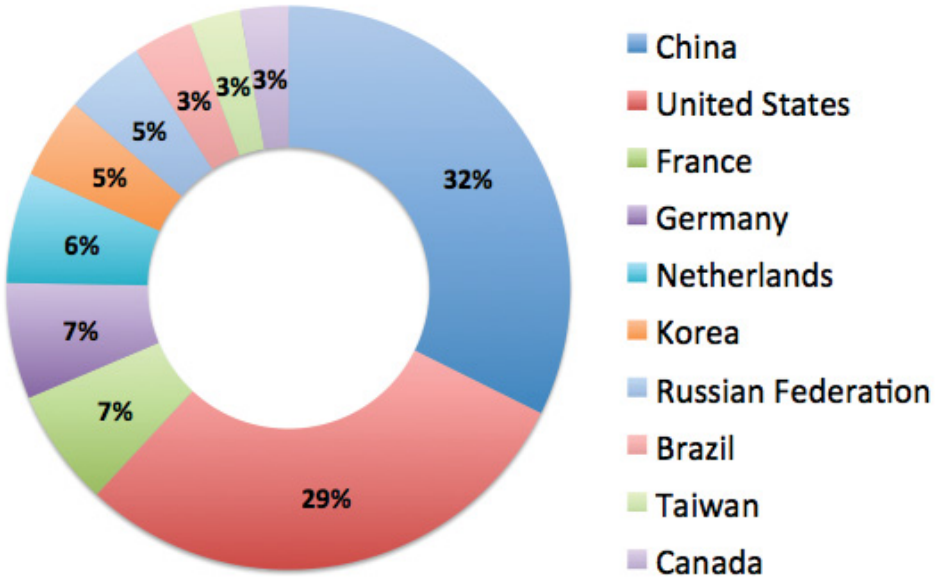


Figure 4.14: Country breakdown for honeypot interaction.

## 4.4 ICS Specific Targeting

Shodan is capable of being used as a passive reconnaissance tool to identify Internet-facing ICS devices. This section reports the results of visual packet inspection

and Snort IDS analysis to identify ICS specific targeting and contrast the rate of targeting pre-identification versus post-identification.

### *4.4.1 Visual Packet Inspection.*

The Allen-Bradley ControlLogix 1756-L61 web server provides an implementation of HTML utilizing Active Server Page (ASP) and Extensible Markup Language (XML) files. To identify reconnaissance activity via web management interface traversal, a visual inspection of the network traffic was conducted. Analysis sought to identify specific HTTP Get requests containing device ASP files coinciding with device traversal and authentication (Table 4.11). Throughout the 55 day deployment analysis did not reveal any device reconnaissance as defined by a web management console device traversal. Reconnaissance is a primary element of network attack and it is assumed ICS specific targeting would include detailed device inspection. In addition, analysis did not reveal any attempts, successful or unsuccessful, to access secured areas of the web management console. Secured areas include Device Configuration, User Management, and Server Management. Regardless of the device configuration, standard or banner mangled, analysis did not reveal any evidence of device specific targeting via connection attempts to the Allen-Bradley ControlLogix management port 44818.

### *4.4.2 Snort IDS.*

To identify and enumerate specific ICS targeting this research utilizes an implementation of the Security Onion, a Linux distribution designed for intrusion detection, network security monitoring, and log management. In addition to the latest Snort implementations, the Digital Bonds Quickdraw SCADA ICS signatures were used which include DNP3, EtherNet/IP, Modbus TCP, and vulnerability signatures. Analysis did not reveal any ICS specific targeting as defined by Digital Bond Quickdraw SCADA IDS signatures. This includes any interaction with known ICS protocols.

Table 4.11: Visual Packet Inspection - device traversal.

| | |
|---|---|
| Homepage | • home.asp |
| Device Diagnostics | • diagover.asp<br>• diagnetwork.asp<br>• msgconnect.asp<br>• etherstats.asp |
| Device Configuration (Secured) | • identity.asp<br>• network.asp<br>• services.asp<br>• emailConfig.asp |
| User Management (Secured) | • editusers.asp<br>• editlimits.asp |
| Server Management (Secured) | • webManage<br>• webTime<br>• backupRestore.html<br>• serverlog.asp |
| Chassis Identification | • chassisWho.asp |

Of over 14 thousand alerts, Snort identified only one high alert. The alert was a scan against the Standard1 honeypot. This scan was categorized as indiscriminate targeting of Internet-facing devices as it used the ZmEu Scanner designed to identify servers with vulnerable versions of PHPMyAdmin and the Allen-Bradley web server uses ASP rather than PHP. Every request for a specific PHP page received a "404 Site or Page not found" response. Identified alerts appear to be generically targeting Internet-facing devices; no alerts specifically targeting ICS devices were identified. Table 4.12 identifies the top 5 alerts identified via Snort analysis accounting for over 96% of all alerts identified. Note these alerts account for non-ICS specific targeting. Each alert indicates the generic alert

title, generator ID, signature ID, count, and a description of the alert. The alerts included scans and automated targeting of Windows 95, Windows 98, and Windows ME machines. All four honeypots received a wide range of scanning activity. These scans registered as medium or low on the Snort priority scale. Scanning tools identified are readily available online (e.g., SIPvicious VOIP scanner) and indicate indiscriminate targeting of web-servers rather than specific device targeting. While these scans do not appear to target ICS specific devices, it reinforces the notion of device vulnerability purely based on Internet-facing deployment. In addition, cyber incidents such as the David-Besse Slammer worm incident have shown the implications of indiscriminate attacks against ICS.

Table 4.12: Description of Snort alerts.

| **Alert (Generator ID, Sig ID)** | **Count** | **Description** |
|---|---|---|
| HTTP_Inspect (120,8) | 3083 | Message WITH INVALID CONTENT-LENGTH OR CHUNK SIZE |
| HTTP_Inspect (120,3) | 5764 | NO CONTENT-LENGTH OR TRANSFER-ENCODING IN HTTP RESPONSE |
| HTTP_Inspect (119,31) | 391 | Unknown Method |
| Stream5 (129,15) | 2982 | Reset outside window |
| Stream5 (129,12) | 1982 | TCP Small Segment Threshold Exceeded |

Although, analysis did not reveal indications of specific ICS targeting, the honeypots did receive Snort alerts relative to apparent non-ICS specific and indiscriminate targeting of Internet-facing devices. A comparative analysis was conducted on total snort alerts to determine if the number of alerts increased post Shodan identification. Figure 4.15 details the total Snort alerts per day for each honeypot. Note these are not representative of ICS

specific targeting and merely provide an indication in the level of interaction post Shodan indexing. All four honeypots indicated a positive linear trend for snort alerts; however, the r-squared values indicate the actual data does not follow this trend. Figure 4.16 details an analysis of mean averages for pre-identification subsets to the mean averages for post-identification subsets using a 95% confidence interval. For each honeypot the post identification traffic does not have a statistically higher mean than the pre-identification averages.

A pairwise t-test is conducted across each data subset to determine if post-identification alert levels are significantly different from pre-identification levels. Table 4.13 through Table 4.16 show the p-values from these results. A p-value of less than 0.05 indicates that the null hypothesis, Shodan does not increase the number of alerts, should be rejected within a 95% confidence interval. These values are indicated in bold within the tables. As the table indicates, in a majority of comparisons (86%) the p-value was over 0.05 indicating no statistical difference in the number of alerts as a result of Shodan identification. In addition, the range of p-values is so varied, there is no definitive indication of an increased number of alerts as a result of Shodan identification.

Table 4.13: Standard1 honeypot Snort IDS alerts - pairwise t-test results.

| | Snort Alerts | | | | |
|---|---|---|---|---|---|
| | Post1 | Post2 | Post3 | Post4 | Post5 |
| Pre2 | 0.06428 | 0.11963 | 0.16389 | 0.05393 | **0.04837** |
| Pre3 | **0.02389** | 0.11054 | 0.11893 | **0.01098** | 0.06069 |

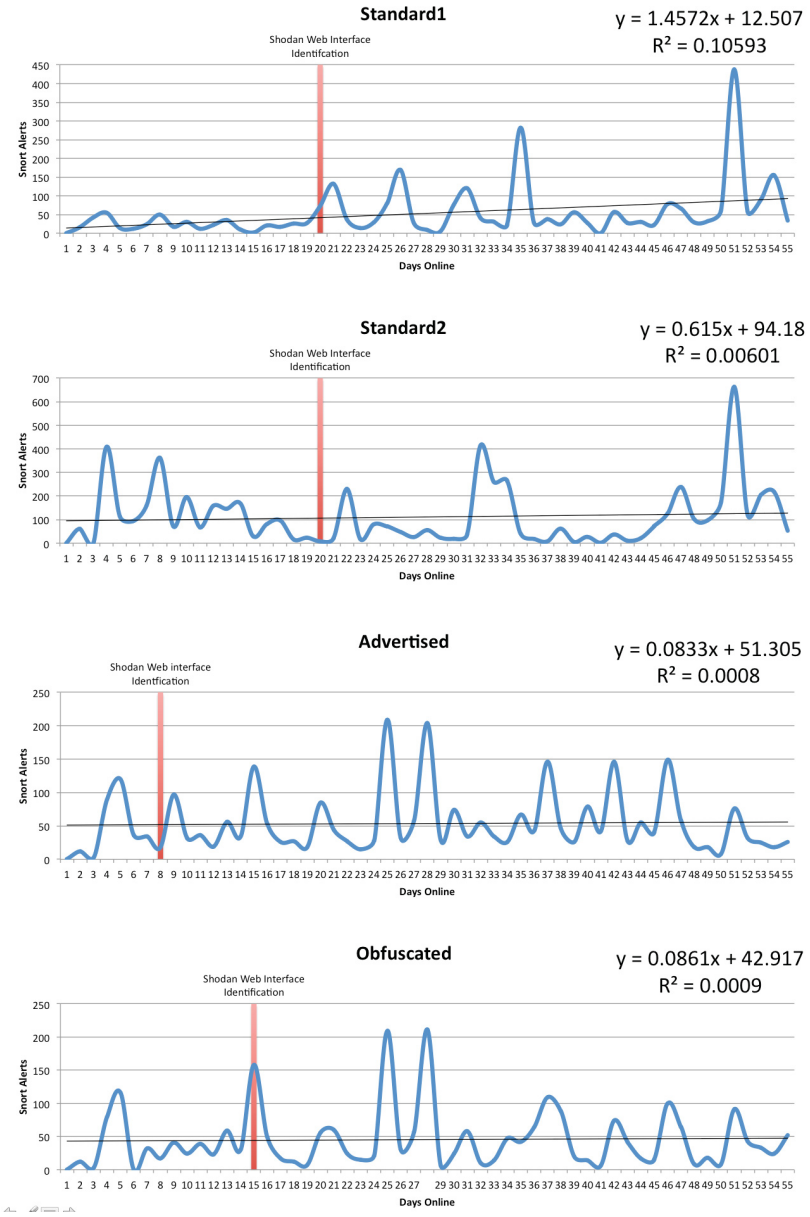


Figure 4.15: Comparative analysis - Linear trending over the 55 day deployment.

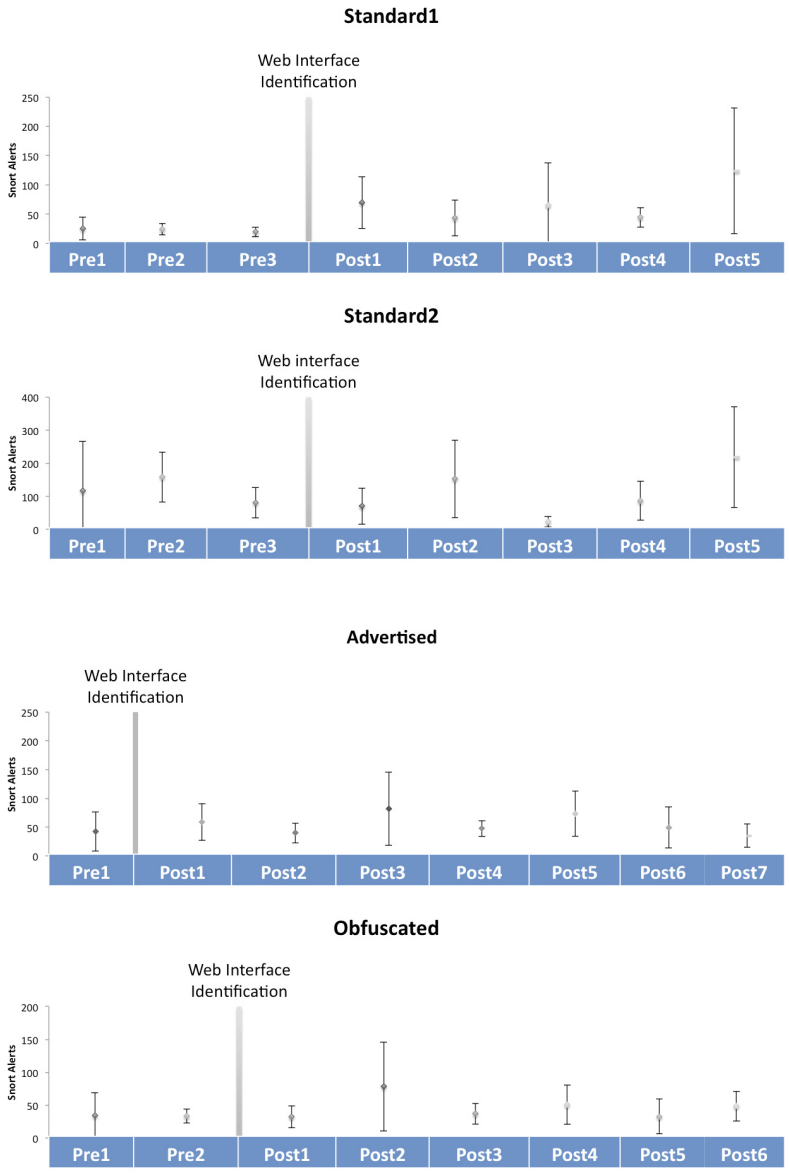


Figure 4.16: Comparative analysis - Subset mean averages pre-identification versus post-identification (95% confidence intervals).

Table 4.14: Standard2 honeypot Snort IDS alerts - pairwise t-test results.

| | Snort Alerts | | | | |
|---|---|---|---|---|---|
| | Post1 | Post2 | Post3 | Post4 | Post5 |
| Pre2 | 0.06639 | 0.47150 | **0.01118** | 0.13077 | 0.13592 |
| Pre3 | 0.33708 | 0.17284 | **0.02209** | 0.44904 | 0.08429 |

Table 4.15: Advertised honeypot Snort IDS alerts - pairwise t-test results.

| | TCP Connections | | | | | |
|---|---|---|---|---|---|---|
| | Post1 | Post2 | Post3 | Post4 | Post5 | Post6 |
| Pre1 | 0.27270 | 0.44723 | 0.17792 | 0.40913 | 0.15963 | 0.40736 |

Table 4.16: Obfuscated honeypot Snort IDS alerts - pairwise t-test results.

| | Snort Alerts | | | | |
|---|---|---|---|---|---|
| | Post1 | Post2 | Post3 | Post4 | Post5 |
| Pre2 | 0.44820 | 0.17835 | 0.45323 | 0.31060 | 0.46933 |
| Pre3 | 0.45464 | 0.10916 | 0.33990 | 0.15361 | 0.48908 |

## 4.5 Banner Impact

As none of the ICS honeypots received targeted ICS specific attacks, it is not possible to measure the impact of banner manipulation relative to attack levels. Despite the lack of targeting, banner manipulation did impact the ability to identify devices via Shodan.

A pilot study was conducted to examine the effects of banner manipulation on Shodan web interface device identification with the goal of assessing the ability to limit device exposure from ICS specific targeting. During the study, an ICS researcher was tasked to use Shodan to identify Allen-Bradley PLCs. Note the researcher was unaware of the service banner modifications made to the mangled honeypots. Each listing of potential devices was examined to determine if the Advertised and Obfuscated devices were exposed based on the search query.

Table 4.17 details the initial round of Shodan search queries using basic knowledge of an Allen-Bradley PLC, to include manufacturer, model, device type, and service ports. Table 4.18 details a second round of search queries with the researcher using specific information regarding the Allen-Bradley ControlLogix PLC port 80 web management service banner. Each table outlines the individual search queries, the total number of devices identified, and an indication of which device was among the query results. Note that the results for Standard1 and Standard2 were the same for each query. For nearly every search query, the Advertised honeypot was among the Shodan results and in multiple cases the Advertised honeypot was the sole device. Alternatively, the Obfuscated honeypot was only identified in two cases, search results containing devices offering port 80 and index.html. With a port 80 query returning 170,467,439 results and index.html returning 1,021,319 results, it is unlikely the Obfuscated honeypot would be directly targeted as an ICS device. The results of a comparison of all four honeypots indicate the Advertised honeypot is likely to be more readily identified for ICS specific targeting based

on the ease of identification using simple search queries, while the Obfuscated honeypot is nearly unidentifiable save for two queries with results extending into the millions.

Table 4.17: Shodan results - basic knowledge of the Allen-Bradley ControlLogix PLC.

| Query | Devices Identified | Standard1/Standard2 | Advertised | Obfuscated |
|---|---|---|---|---|
| Allen | 3350 | | X | |
| Bradley | 2638 | | X | |
| Allen-Bradley | 98 | | X | |
| ControlLogix | 1 | | X | |
| PLC | 11958 | | X | |
| Port:80 | 170467439 | X | X | X |
| Port:44818 | 0 | | | |
| Allen ControlLogix | 1 | | X | |
| Allen PLC | 8 | | | |
| Allen port:80 | 83 | | X | |
| Allen-Bradley ControlLogix | 1 | | X | |
| Allen-Bradley PLC | 7 | | | |
| Allen-Bradley port:80 | 1 | | X | |
| Allen-Bradley PLC port:80 | 0 | | | |
| Bradley ControlLogix | 1 | | X | |
| PLC ControlLogix | 0 | | | |

## 4.6 Discussion

This research provides an extension of previous work in the ICS honeypot arena. Wilhoit's research found a significant level of device targeting, noting attacks within 18 hours of deployment and 39 targeted attacks over the 28 day deployment [52]. In addition, Wilhoit's subsequent ICS honeynet deployment resulted in 74 attacks, 11 defined as critical, over a 90 day deployment. Wilhoit defined critical as an attack with unestablished motivation but capable of catastrophic failure of an ICS device's operations [52]. The

Table 4.18: Shodan results - knowledge of the Allen-Bradley ControlLogix PLC service banner.

| Query | Devices Identified | Standard1/Standard2 | Advertised | Obfuscated |
|---|---|---|---|---|
| GoAhead-Webs | 23428 | X | X | |
| Connection: Close | 102449800 | X | X | |
| index.html | 1021319 | X | X | X |
| GoAhead-Webs Connection: Close | 51104 | X | X | |
| GoAhead-Webs index.html | 16184 | X | X | |
| Connection: Close index.html | 338683 | X | X | |
| GoAhead-Webs Connection: Close index.html | 490 | X | X | |

remaining 63 attacks were classified as non-critical, as defined as the inability to cause catastrophic failure (e.g., denial of service attack) [52]. Wilhoit also noted 33,466 automated attacks originating from 1,212 unique IP addresses. Wilhoit's results appear inconsistent with the findings of this research, wherein no targeted ICS attacks were identified. The primary differences between the two research efforts include: specific research goals and individual honeypot implementation.

Wilhoit's primary objective was to assess who is attacking Internet-facing ICS devices and provide indications as to attack motivations. As such, Wilhoit intentionally took steps to solicit the honeypots, seeding the devices on Google, Pastebin, and Shodan, while also utilizing naming conventions to readily identify the devices (e.g., SCADA-1). Alternatively, the primary goal of this research was to evaluate the impact of Shodan on Internet-facing ICS devices, utilizing a generic characterization of ICS devices currently identifiable via Shodan for honeypot design and configuration. As such, devices were deployed unsolicited and utilizing naming conventions akin to Allen-Bradley ControlLogix PLCs currently identifiable via Shodan (i.e., a descriptive device name incorporating device deployment location). Each honeypot device was configured with a device name designed to infer the device was newly deployed and associated with a water

utility (e.g., ab.2013.water.s3). The vast contrast in research findings may be a result of device solicitation, indicating malicious actors are utilizing other avenues of ICS device reconnaissance beyond Shodan.

Wilhoit utilized both high-interaction and low-interaction honeypots. Wilhoit's high-interaction honeypots were comprised of a simulated PLC implemented via Honeyd and a HMI implemented via a Dell DL360 server, each acting as a basic PHP web server. Wilhoit's results identified multiple attacks in the form of targeted PHP attacks against individual honeypots. Alternatively, the honeypots deployed in support of this research utilized physical Allen-Bradley ControlLogix PLCs whose web server uses an implementation of XML and ASP. Despite the lack of PHP, analysis identified a number of apparent automated attacks targeting PHP web servers. It is possible the PHP targeted attacks identified by Wilhoit were due to the particular implementation of an Internet-facing PHP web server, rather than specific ICS targeted attacks. In addition, it is possible the existence of potentially vulnerable PHP web servers increased device identification and subsequent targeting.

In addition to specific web server implementation, Wilhoit's simulated PLC and HMI honeypots lacked functional and operating characteristics associated with actual PLC devices. Indeed, the PLC honeypot used an implementation of Honeyd, utilizing python to present a generic web page simulating a water pressure station. Malicious device modifications and user reads/writes could be accomplished by changing values in a web form. The HMI honeypot offered attackers direct interaction with a simulated HMI wherein any simple modification (e.g., clicking to open a valve) was deemed a targeted attack against the associated PLC. Alternatively, the physical PLCs utilized in support of this research required actual authentication, albeit default username and password, to perform any modification via the web management console. In addition, to alter PLC operations required direct interaction with the device EtherNet/IP management service, as

well as additional software to upload, download, and alter ladder logic. Note that past research has indicated that use of actual devices for honeypots, as in this research, provides a better indicator of attack tactics than simulated devices, as in Wilhoit's research [44].

## 4.7 Summary

This chapter provides results of network analysis performed for this research. An examination of Shodan's indexing routine revealed all four newly deployed and unsolicited devices were successfully indexed and identified within 19 days. Analysis of network activity post Shodan identification provided no indication of increased device interaction, as defined by TCP connections, total TCP packet count, and unique IP addresses. Neither visual packet inspection nor Snort IDS analysis revealed any instances of ICS specific device targeting or attacks, as defined by web management traversal, attempted access to secured areas of the web management interface, port 44818 interrogation and interaction. Given no ICS specific attacks, a comparison of the ability to limit device targeting via banner manipulation was not possible. However, pilot study results indicate service banner manipulation decreases device susceptibility to identification vis Shodan search engine queries. Findings indicate Shodan is a capable ICS reconnaissance tool, with the ability to index and identify unsolicited Internet-facing ICS devices. The results of this research, however, indicate Shodan does not currently impact Internet-facing ICS device security, but it is expected Shodan will become a more prevalent passive reconnaissance tool based on its capability to readily identify ICS devices.

# V. Conclusions

Cyber attacks akin to Stuxnet and the Slammer worm illustrate the potential exploitation of ICS via both targeted and inadvertent attacks, each capable of significant damage. Throughout the ICS security community the vulnerability and fragility of ICS is widely understood and the once mythical ICS air gap was shattered by the work of Leverett and Project SHINE. Shodan is a capable passive reconnaissance tool, whose scanning routine ensures most Internet-facing devices will be indexed given enough time. Additionally, a basic knowledge of devices and services allows users to craft device-specific signatures and create targeted lists of Internet accessible devices.

Findings indicate network activity did not increase post indexing in Shodan, and most targeting appeared indiscriminate. Previous work by Wilhoit with ICS honeypots indicated a number of targeted attacks; however, this research did not collaborate these findings as no ICS-specific targeted attacks were observed. This is likely based on the difference in honeypot design, implementation, and solicitation.

## 5.1 Conclusions

The exponential growth of the Internet, increased network connectivity, and the need for remote access has led to a significant number of ICS devices directly connected to the Internet. Previous research efforts revealed an exorbitant number of ICS devices are currently deployed Internet-facing [6, 30] and Shodan is able to readily identify these devices. Security professionals and news outlets at large publicize the perils of Shodan and by all indications Shodan should be categorized as a threat to Internet-facing ICS, however, there lacks empirical evidence linking Internet-facing device targeting to Shodan device identification.

The overall goal of this research is to evaluate Shodan's impact on Internet-facing ICS device security by deploying a series of high-interaction ICS honeypots. The primary goals are to evaluate Shodan indexing functionality, contrast network activity levels as a result of Shodan identification, and enumerate any ICS specific targeting and attacks. An examination of Shodan's indexing routine revealed all four newly deployed and unsolicited devices were successfully indexed and identified within 19 days. Analysis of network activity post Shodan identification provided no indication of increased network activity. Neither visual packet inspection nor Snort IDS analysis revealed any instances of ICS specific device targeting or attacks.

The secondary goal of this research is to assess the impact of ICS device service banner data relative to device identification within Shodan and subsequently evaluate the ability to limit Shodan device exposure via banner manipulation. The lack of specific ICS targeting prohibited an evaluation of the ability to limit device targeting via banner manipulation was not possible. However, pilot study results indicate service banner manipulation decreases device susceptibility to identification vis Shodan search engine queries.

The overall findings indicate Shodan does not currently impact Internet-facing ICS device security, but research has demonstrated Shodan's utility as a passive reconnaissance tool. With the continued growth and connectivity of ICS devices, it is expected Shodan will become more commonly used tool to target ICSs.

## 5.2 Future Work

This research presented an evaluation of Shodan's impact on Internet-facing ICS device security. This section presents ideas for future work regarding the ICS honeypots.

### *5.2.1 Deployment Location.*

The ability to obtain Internet-facing IP space co-located with an ICS entity limited the honeynet design and the size of the honeynet. Available resources allowed for the

deployment of four ICS honeypots and dictated the honeypots be deployed with sequential static IP addresses in the same subnet. Future research should seek a broader deployment across multiple venues to include: commercial IP space, residential IP space, and co-located with government networks (e.g., outside Wright-Patterson .mil network). In addition, future work should look to deploy honeypots in multiple critical infrastructure sectors to include oil and gas, water distribution systems, and electrical utilities.

#### *5.2.2 Deployment Length.*

The deployment period for this research was 55 days based on previous ICS honeypot research and an approximation of the time required to scan all public IPv4 addresses. Future research should extend honeypot deployment to six months or as long as a year to allow for a larger dataset for comparative analysis and trending.

#### *5.2.3 Honeypot Type.*

Future research should consider the utilization of low-interaction honeypots alongside high-interaction honeypots. Low-interaction honeypots would open the Amazon EC2 cloud deployment venue, offering the ability to deploy multiple ICS honeypots creating a larger dataset for analysis. In addition, Honeyd offers the ability to proxy requests to any predefined IP address. As such, future work could use an Amazon Cloud deployment with Honeyd proxying port 80 and port 44818 requests to a physical PLC, creating a hybrid ICS honeypot offering the advantages of both high and low-interaction honeypots.

#### *5.2.4 Honeypot Design.*

The honeypots deployed in support of this research were designed and configured to be representative of ICS devices currently identifiable via Shodan. Configuration also utilized default username and password to simulate a newly deployed PLC. To supplement the notion of a newly deployed PLC, devices were deployed with the default system time. For ControlLogix system time defaults to 1 January 1970. As presented in

this research, Shodan queries are able to use this system time to identify devices which have been online for a specific time period. Future research should consider changing the system time on devices to represent a device which has been online for multiple years, potentially providing a more enticing target for exploitation.

#### *5.2.5 Programmable Logic Controller.*

This research uses the Allen-Bradley ControlLogix 1756-L61 CPU module and eWeb Ethernet module. Future work should incorporate both additional manufacturers and models. At a minimum, device manufacturers should be extended to include Siemens and Schneider as they represent the number two and three primary PLC suppliers in North America [2].

#### *5.2.6 Shodan Device Categorization.*

Shodan is a capable passive reconnaissance tool able to readily identify Internet-facing ICS devices. Utilizing only three key terms relating to the Allen-Bradley ControlLogix PLC, over 490 devices were identified. Future research should investigate and classify ICS devices currently identifiable via Shodan relative to United States critical infrastructure sectors.

### 5.3 Concluding Remarks

The protection of United States critical infrastructure is vital to national security and ICSs are the backbone of many critical infrastructure sectors. Trends indicate ICS growth is expected to flourish and the innate demand for availability has resulted in an increased level of Internet connectivity. Homeland security officials have warned that the obscurity that previously protected many industrial control systems is quickly disappearing in a flood of digital light [39]. In 2009, the Shodan computer search engine was launched creating a database of Internet-facing devices, identifying hundreds of millions of devices over the past four years, most notably an untold number of ICS devices. Shodan is capable of functioning as a passive reconnaissance tool capable of specifically identifying ICS

devices exposed to the Internet. Independent research definitively proved Shodans capabilities, identifying thousands of ICS associated devices, many with weak or default authentication. Subsequent research provided evidence of malicious actors directly attacking a simulated United States water control system. This research sought to correlate Shodan device identification with direct ICS targeting and measure Shodans impact on Internet-facing ICS device security. Findings indicate although Shodans scanning routine virtually guarantees the eventual indexing and identification of Internet-facing ICS devices, a measurement of network activity post identification does not indicate Shodan is actively being used as a reconnaissance tool for ICS attack. Despite these findings, Shodan is a more than capable of identifying exposed ICS devices and as such poses a real threat to Internet-facing ICS and thereby national security. It is expected as ICS vulnerabilities and exploits become more readily known, Shodans will become a primary tool for malicious actors to directly target ICSs.

## Bibliography


[1] Abrams, M. and J. Weiss. "Malicious Control System Cyber Security Attack Case Study Maroochy Water Services, Australia". *The MITRE Corporation*, August 2008.

[2] Automationprimer.com. "PLC Manufacturer Rankings", October 2013. URL http://automationprimer.com/2013/10/06/plc-manufacturer-rankings/.

[3] Boyer, S. *SCADA: supervisory control and data acquisition*. International Society of Automation, Reading MA, 2009.

[4] Burk, D. "Security Onion - intrusion detection, network security monitoring, and log management", February 13 2014. URL http://blog.securityonion.net/.

[5] Butts, J. *Storming the Castle: Pitfalls of Defense-in-Depth Strategies*. 2013. URL http://www.sans.org/event/north-american-scada-2013.

[6] Byres, E. "Project SHINE: 1,000,000 Internet-Connected SCADA and ICS Systems and Counting", September 19 2013. URL http://www.tofinosecurity.com/blog/project-shine-1000000-internet-connected-scada-and-ics-systems-and-counting.

[7] Byres, E., D. Leversage, and N. Kube. "Security incidents and trends in SCADA and process industries". *The Industrial Ethernet Book*, 39(2):12–20, 2007.

[8] Cardenas, A., S. Amin, and S. Sastry. "Research Challenges for the Security of Control Systems". *Hotsec08 - 3rd USENIX Workshop on Hot Topics in Security*. July 2008.

[9] Claburn, T. "CIA Admits Cyberattacks Blacked Out Cities", January 2008. URL http://www.informationweek.com/cia-admits-cyberattacks-blacked-out-cities/d/d-id/1063513?

[10] Clarke, G., D. Reynders, and E. Wright. *Practical modern SCADA protocols: DNP3, 60870.5 and related systems*. Newnes, Burlington MA, 2004.

[11] ControlSys.org. "The Control System Integrators Association (CSIA)", 2014. URL www.controlsys.org.

[12] Department of Homeland Security. "ICS-CERT Alerts", January 2014. URL http://ics-cert.us-cert.gov/alerts/.

[13] Digital Bond SCADA Security Portal. "Digital Bond:SCADA Honeynet", 2013. URL http://www.digitalbond.com/tools/scada-honeynet/.

[14] DNP.org. "DNP3:A Distributed Network Protocol FAQ", March 2013. URL http://www.dnp.org/default.aspx.

[15] Even, L. “SANS Intrusion Detection FAQ: What is a Honeypot?”, July 2000. URL http://www.sans.org/security-resources/idfaq/honeypot3.php.

[16] Franz, M. and V. Pothamsetty. “SCADA HoneyNet Project: Building Honeypots for Industrial Networks”, July 15 2005. URL http://scadahoneynet.sourceforge.net/.

[17] Gasper, P. “Cyber Threat to Critical Infrastructure 2010-2015:Increased Control System Exposure”. September 2008. URL http://usacac.army.mil/cac2/cew/repository/presentations/15_Idaho_Natl_Lab_IACS-CI_Threat_2010-2015.pdf.

[18] Goldman, D. “Shodan: The scariest search engine on the Internet”, April 2013. URL http://money.cnn.com/2013/04/08/technology/security/shodan/.

[19] Government Accounting Office. *Challenges and Efforts to Secure Control Systems*. Technical Report GAO-04-354, Washington DC, March 2004. URL http://www.gao.gov/new.items/d04354.pdf.

[20] Grimes, R. *Honeypots for Windows: A Honeypot Deployment Plan*. Apress, New York NY, 2005.

[21] Hacker, T., B. Noble, and B. Athey. “The effects of systemic packet loss on aggregate TCP flows”. *Supercomputing, ACM/IEEE 2002 Conference*. IEEE, November 2002.

[22] Higgins, K. “The SCADA Patch Problem”, January 2013. URL http://www.darkreading.com/vulnerability/the-scada-patch-problem/240146355.

[23] Hildick-Smith, A. “Security for Critical Infrastructure SCADA Systems”. *SANS Reading Room*, February 2005.

[24] Hill, D., D. Woll, and P. MIller. “Process control in the HPI: A not-so-sentimental journey”, January 2012. URL http://www.hydrocarbonprocessing.com/Article/3050923/Process-control-in-the-HPI-A-not-so-sentimental-journey.html.

[25] HoneyNet.org. “HoneyNet: Conpot”, May 2013. URL http://www.honeynet.org/node/1047.

[26] Hoover, N. “Thousands Of Industrial Control Systems At Risk: DHS Study”. *InformationWeek*. January 2013. URL http://www.informationweek.com/security/risk-management/thousands-of-industrial-control-systems--at-risk-dhs-study/d/d-id/1108149?

[27] Keefe, M. “Timeline: Critical infrastructure attacks increase steadily in past decade”. *Computer World*. November 2012. URL http://www.computerworld.com/s/article/9233173/Timeline_Critical_infrastructure_attacks_increase_steadily_in_past_decade.

[28] Krutz, R. *Securing SCADA systems*. John Wiley and Sons, Indianapolis IN, 2005.

[29] Kumar, M. “Chinese hackers APT1 honeypot water control system”. *The Hacker News*. August 2013. URL http://thehackernews.com/2013/08/Chinese-hackers-APT1-honeypot-water-control-system.html.

[30] Leverett, E. *Quantitatively Assessing and Visualising Industrial System Attack Surfaces*. Ph.D. thesis, University of Cambridge, United Kingdom, June 2011. URL https://www.cl.cam.ac.uk/~fms27/papers/2011-Leverett-industrial.pdf.

[31] Lewis, T. *Critical infrastructure protection in homeland security: defending a networked nation*. John Wiley and Sons, Hoboken NJ, 2006.

[32] Maxmind.com. “MaxMind - GeoIp”, February 2014. URL http://www.maxmind.com/en/geolocation_landing.

[33] Meserve, J. “Sources: Staged cyber attack reveals vulnerability in power grid”. *CNN*. September 2007. URL http://www.cnn.com/2007/US/09/26/power.at.risk/index.html.

[34] Mitchell, R. “After Stuxnet: The New Rules of Cyberwar”. *Computerworld*. November 2012. URL http://www.computerworld.com/s/article/9233158/After_Stuxnet_The_new_rules_of_cyberwar.

[35] MODBUS.org. “MODBUS FAQ”, 2013. URL http://www.modbus.org/faq.php.

[36] Moteff, J. and P. Parfomak. “Critical infrastructure and key assets: definition and identification”. *CRS Report for Congress*. DTIC Document, Washington DC, 2004.

[37] Networecon.com. “Online Internet Scanning Calculator”, December 2012. URL http://www.networecon.com/tools/scancalc/#.UwanW2RDvqE.

[38] ODVA. “Open DeviceNet Vendors Association (ODVA) EtherNet/IP (Ethernet Industrial Protocol)”, 2013. URL http://www.odva.org/default.aspx?tabid=67.

[39] O‘Harrow, R. “Cyber search engine Shodan exposes industrial control systems to new risks”. *The Washington Post*. June 2012. URL http://www.washingtonpost.com/investigations/cyber-search-engine-exposes-vulnerabilities/2012/06/03/gJQAIK9KCV_story.html.

[40] Ramsey, F. and D. Schafer. *The Statistical Sleuth: A Course in Methods of Data Analysis: A Course in Methods of Data Analysis*. Cengage Learning, Boston MA, 2012.

[41] Rrushi, J. *SCADA protocol vulnerabilities*, 150–176. Critical Infrastructure Protection. Springer Berlin Heidelberg, New York NY, 2012.

[42] SANS Institute. “SANS Penetration Testing”, February 2014. URL http://pen-testing.sans.org/.

[43] Shodan. “SHODAN Computer Search”, October 2013. URL www.shodanhq.com.

[44] Spitzner, L. *Honeypots: tracking hackers*. Addison-Wesley Reading, Reading, MA, 2003.

[45] Stouffer, K., Joe Falco, and Karen Scarfone. “Guide to industrial control systems (ICS) security”. *NIST Special Publication*, 800(82):16–16, 2008.

[46] Symantec. “SCADA (Supervisory Control and Data Acquisition) security threat landscape”, 2012. URL http://www.symantec.com/threatreport/topic.jsp?aid=scada_vulnerabilities&id=vulnerability_trends.

[47] TOFINO Security. “Davis-Besse: Cyber Incident Case Profile CP-103v1”, January 2009. URL http://www.tofinosecurity.com/sites/default/files/CP-103-Case_Profile-Davis_Besse-rev1.pdf.

[48] U.S. Department of Homeland Security. *CSAR-10-025-01 Analysis of Shodan Computer Search Engine*. Technical Report CSAR-10-025-01, Department of Homeland Security, October 2010. URL https://portal.us-cert.gov/member/libraryV3/rhsIndex.cfm?action=9&returnAction=32&libid=364780.

[49] U.S. Department of Homeland Security. “DHS Common Cybersecurity Vulnerabilities ICS 2010”, September 2010. URL https://ics-cert.us-cert.gov/pdf/DHS_Common_Cybersecurity_Vulnerabilities_ICS_2010.pdf.

[50] U.S. Department of Homeland Security. *ICS-CERT Monitor (October 2012)*. Technical Report ICS-MM201210, United States Computer Emergency Readiness Team, October 2012. URL http://ics-cert.us-cert.gov/sites/default/files/ICS-CERT_Monthly_Monitor_Oct-Dec2012_2.pdf.

[51] Wade, S. *SCADA Honeynets: The attractiveness of honeypots as critical infrastructure security tools for the detection and analysis of advanced threats*. Ph.D. thesis, Iowa State University, Ames IA, 2011. URL http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=3130&context=etd.

[52] Wilhoit, Kyle. “The SCADA That Didn’t Cry Wolf”. *Trend Micro Incorporated*, March 2013.

[53] Wilhoit, Kyle. “Who’s Really Attacking Your ICS Equipment?” *Trend Micro Incorporated*, August 2013.

[54] WindmillSoft. “Windmill Soft HMI Software - Data acquisition and control”, January 2013. URL http://www.windmillsoft.com/daqshop/hmi-software.html.

[55] Zhu, B., A. Joseph, and S. Sastry. “A taxonomy of cyber attacks on SCADA systems”. *2011 International Conference on Cyber, Physical and Social Computing*, 380–388. IEEE, October 2011.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704–0188*



| 1. REPORT DATE *(DD–MM–YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From — To)* |
|---|---|---|
| 27–03–2014 | Master's Thesis | Oct 2012–Mar 2014 |

**4. TITLE AND SUBTITLE**

Impact of the Shodan Computer Search Engine on Internet-facing industrial control system devices

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Bodenheim, Roland C., Captain, USAF

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB, OH 45433-7765

**8. PERFORMING ORGANIZATION REPORT NUMBER**

AFIT-ENG-14-M-14

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Department of Homeland Security ICS-CERT
POC: Nick Carr
245 Murray Lane SW Bldg 410, Mail Stop 635
Washington, DC 20528
Nicholas.Carr@hq.dhs.gov

**10. SPONSOR/MONITOR'S ACRONYM(S)**

DHS ICS-CERT

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

**13. SUPPLEMENTARY NOTES**



**14. ABSTRACT**

The Shodan computer search engine crawls the Internet attempting to identify any connected device. Using Shodan, researchers identified thousands of Internet-facing devices associated with industrial controls systems (ICS). This research examines the impact of Shodan on ICS security, evaluating Shodans ability to identify Internet-connected ICS devices and assess if targeted attacks occur as a result of Shodan identification. In addition, this research evaluates the ability to limit device exposure to Shodan through service banner manipulation. Shodans impact was evaluated by deploying four high-interaction, unsolicited honeypots over a 55 day period, each configured to represent Allen-Bradley programmable logic controllers (PLC). All four honeypots were successfully indexed and identifiable via the Shodan web interface in less than 19 days. Despite being indexed, there was no increased network activity or targeted ICS attacks. Although results indicate Shodan is an effective reconnaissance tool, results contrast claims of its use to broadly identify and target Internet-facing ICS devices. Additionally, the service banner for two PLCs were modified to evaluate the impact on Shodan indexing capabilities. Findings demonstrated service banner manipulation successfully limited device exposure from Shodan queries.

**15. SUBJECT TERMS**

Shodan, ICS, PLC, critical infrastructure protection

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Maj Jonathan W. Butts, PHD |
| U | U | U | UU | 121 | 19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636 x4332 jonathan.butts@afit.edu |